不道德行为中道德标准对自我欺骗的影响：来自 ERP 的证据*

范伟[1,2,3] 杨颖[1,2] 郭希亚[1,2] 林卓铭[1,2] 钟毅平[1,2,3]

（[1]湖南师范大学教育科学学院心理系，[2]认知与人类行为湖南省重点实验室，长沙，410081）

（[3]湖南师范大学交叉科学研究院，长沙，410081）

摘要 本研究旨在通过事件相关电位技术探讨不道德行为中自我欺骗的心理作用及其神经机制，特别是探究道德标准对自我欺骗的抑制作用。实验 1 考察不道德行为中自我欺骗的内在神经机制。实验 1 通过发送者-接受者范式诱发被试的不道德行为，并通过个体对未知随机概率值的预测来测量自我欺骗。行为结果发现，在欺骗试次中，被试选择低于真实信念的比例显著高于在诚实试次中的比例。脑电结果发现，相比于诚实试次，被试在欺骗试次中会诱发更大的 P2、N2 以及 P300 成分。实验 2 通过道德标准启动任务，探讨对道德标准的关注如何影响自我欺骗。行为结果显示，在控制条件下，被试在欺骗试次中选择低于真实信念的比例显著大于诚实试次中诚实试次中的比例。脑电结果显示，在道德标准启动条件下，欺骗试次诱发的 P2 和 N2 成分显著低于诚实试次。这些研究结果可能表明在不道德行为中，个体易于形成虚假信念导致自我欺骗，而关注道德标准能有效抑制此类现象的发生。

关键词 自我欺骗，不道德行为，道德标准，信念，事件相关电位

分类号 B849; C91; B845

1 引言

自我欺骗(self-deception)，也可以被称之为“自欺”，是指一种有意识的动机性虚假信念，这种虚假信念是与真实信念相矛盾的(Pinker, 2011)。个体有时会出于多种动机，主动扭曲事实，以相信与真实信念相反的虚假信念，即便这些信念不符合客观事实，这种现象即为自我欺骗(Trivers, 2000; 鞠实儿 等, 2003; 钟罗金, 莫雷, 2019)。自我欺骗是一种复杂且普遍存在的心理现象，伴随着情绪和认知的交互作用，几乎与生俱来地存在于个体之中(Foster & Frijters, 2014; Surbey, 2011; Trivers, 2000)。尽管以往研究多聚焦于自我欺骗的积极作用，然

---

*收稿日期：2024-09-10

国家自然科学基金面上项目 (32371126)，湖南省研究生科研创新项目 (CX20230487)。

范伟和杨颖为共同第一作者，对本文贡献等同。

通讯作者：钟毅平，E-mail: ypzhong@hunnu.edu.cn

而在个体心理健康、行为模式及社会层面可能带来的消极影响不容忽视，尤其是在道德领域(Hirschfeld et al., 2008; Johnson, 1995; Lee & Klein, 2002; Martínez-González et al., 2016)。自我欺骗常被个体用作调节自身利益与道德标准冲突的策略，进而维持个人形象，同时也被认为是道德推脱背后的关键心理机制(Batson, 1999; Tang et al., 2018)。研究表明，自我欺骗在不道德行为中的整个过程（行为前预测、行为中决策、行为后回忆与解释）均发挥了重要作用(Epley & Dunning, 2000; Mitchell et al., 1997)。首先，由于人们对自身的固有偏见，绝大多数人倾向于认为自己是道德的(McGregor, 2006; McGregor et al., 2001)，从而高估自己未来从事符合社会期望行为的可能性(Epley & Dunning, 2000)。其次，自我欺骗在即时的不道德决策中也起到关键作用。Tenbrunsel 和 Messick(2004)的研究发现，自我欺骗会导致行为决策出现偏差，使个体能够在不损害自我形象的情况下获取更多个人利益，当个体能够说服自己其行为是符合道德时，判断偏差便会产生。再次，当评估过去不道德的行为时，个体倾向于认为自己的行为比实际更为道德(Mitchell et al., 1997)。这种机制使人们能够为自身行为辩护，并将表面上错误的行为合理化。自我欺骗所导致的大脑“短路”可以颠覆人们的道德信仰，削弱对行为准则和道德规范的重视，成为个人道德发展的重大障碍(Levy, 2004; Turner, 1975)。由于自我欺骗在不道德行为中的普遍存在，这一倾向不仅导致社会中不道德行为的频繁发生，还可能引发种族屠杀和破坏全球合作社会等极为严重的后果(Babino et al., 2018; Jones, 1991; Kish-Gephart et al., 2010)。因此，考察不道德行为中自我欺骗的作用，并深入探讨关注道德标准对自我欺骗的抑制作用具有重要的社会效益。

自我欺骗是一把“双刃剑”，它可以激发积极信念，提升自信心和主观幸福感，促进乐观态度和心理健康(Epley & Whitchurch, 2008; Pears & Pugmire, 1982; Pinker, 2011)。然而，自我欺骗也可能因信息真实性的丧失带来巨大的利益损失，其消极作用尤为显著地体现在道德层面(Von Hippel & Trivers, 2011)。个体在实施不道德行为的同时，可以通过自我欺骗维持对自身道德形象的信念(Moore, 2016; Tenbrunsel & Messick, 2004)。因此，自我欺骗是道德衰退的催化剂，助长了不道德行为的发生(Levy, 2004; Lu & Chang, 2011)。

近年来，越来越多的研究者开始关注自我欺骗在不道德行为中的作用(Levy, 2004; Lu & Chang, 2011; Rick et al., 2008; Tenbrunsel et al., 2010; Tenbrunsel & Messick, 2004)。自我欺骗被认为是道德推脱背后的关键心理机制，个体通过自我欺骗来合理化不道德行为，从而维护自身的道德形象(Bandura et al., 1996; Batson, 1999; Tang et al., 2018)。大多数人普遍认为自己具有道德品质(Greenwald, 1980; Von Hippel & Trivers, 2011)，并且倾向于夸大那些被社会高度重视的积极特质，特别是在道德形象方面(Chance & Norton, 2015; Shu & Gino, 2012)。因

此，当个体采取与道德标准相悖的行为时，通常会经历心理上的内部冲突与认知失调(Festinger & Freedman, 1964)。作为一种防御机制，自我欺骗能够通过调整自身信念，将不道德行为在认知上重新定义为合理或正当的，从而减少心理冲突，维护道德形象(Trivers, 2000; Turk, 2012)。研究发现个体在实施不道德行为后，往往通过自欺重新调整自己的道德认同，以缓解内心的不和谐感。自我欺骗使得个体可以将本质上不道德的行为在认知上视为无道德意义甚至是正当的，进而在实施这些行为时体验到更少的内部冲突(Bandura, 2011; Ditto & Lopez, 1992; Kunda, 1990; Roeser et al., 2016)。

Tenbrunsel 和 Messick(2004)提出了“道德褪色”理论，用以解释自我欺骗在不道德行为中的作用。自我欺骗能够使道德色彩在决策过程中逐渐褪去，将原本具有道德含义的行为重新编码为不涉及道德的行为。在这一过程中，利己主义与道德原则之间的权衡被模糊化，导致行为决策的伦理逐渐减弱，使道德含义变得模糊不清(Kunda, 1990; Roeser et al., 2016; Tenbrunsel & Messick, 2004)。自我欺骗导致对某些信息的忽视和错误信念，这种情况可能会在道德责任的判断和对行为后果的估计中引发严重的错误(Bok, 1989)。自我欺骗就像一种“道德漂白剂”，消除了决策中的道德色彩，使个体能合理化自身的不道德行为(Ditto & Lopez, 1992)。因此，个体在做出行为决策时能够巧妙地绕过自身的道德标准，从而增加不道德行为的发生可能性(Tenbrunsel & Messick, 2004)。此外，道德褪色理论还指出，个体会采取自我欺骗策略积极寻找或创造有利于自我道德形象的证据(Moore, 2016)。个体的信念也会随着内部心理状态和外部情境的变化而不断调整(Johnson & Fowler, 2011)。当个体发现自身的原有观念与新的环境产生冲突时，会调整或改变信念，以适应新的情境。同样地，当个体内部心理状态出现不和谐时，为了恢复心理平衡，个体也会调整信念以维护内部的和谐(Trivers, 2000; Turk, 2012)。这一理论认为，自我欺骗的本质在于不同状态之间的相互转化。然而，在信念调整的过程中，个体并非完全颠覆原有观念，而是将新的信息与已有信念整合，形成更具适应性的信念体系(Politzer & Carles, 2001)。然而，信念调整的相关理论仍然停留在概念层面，缺乏实证研究来验证不道德行为中的自我欺骗过程以及道德褪色现象。

另外，自我欺骗作为道德推脱背后的内在心理机制，常被个体用作策略，以处理个人利益与道德标准之间的冲突，从而维持自身的道德形象(Bandura et al., 1996; Batson, 1999; Tang et al., 2018)。通过扭曲自己的道德信念，个体能够进行原本被认为是不道德的行为，并在行为发生后依然保持自认为良好的道德形象(Johnson, 1995; Rick et al., 2008)。已有研究表明，实施暴力行为的人并不一定是由临床病理所驱动的。相反，他们并不为自己的行为带来的道德后果感到担忧，而是通过自我欺骗维持积极的道德自我概念(Shaw et al., 2011)。Tenbrunsel

和 Messick (2004)指出，正是因为自我欺骗能够让人们在不感到内疚的情况下进行不道德行为，它犹如“催化剂”，助长了不道德行为的产生。基于此，本研究提出假设 1：不道德行为更容易诱发自我欺骗行为。

尽管直接探讨道德标准对自我欺骗影响的研究较少，但有学者指出，自我欺骗与道德标准在维持和更新自我概念以及道德推脱中的作用截然相反(Bandura et al., 1996; Fleeson, 2001)。一方面，自我概念维持理论认为，为了维持积极的自我形象，个体的行为会反映在其自我概念中。如果行为不符合社会规范，个体通常会更新自我概念。然而，如果某些不道德行为在自我概念中被合理化为可以接受或不严重的，个体可能不会更新自我概念，进而导致更多不道德行为的发生(Fleeson, 2001; Mischel, 1999)。自我欺骗在不道德行为中往往通过委婉和宽松的方式对自身行为进行定义，将本质上不道德的行为合理化为可接受的(Ditto & Lopez, 1992; Kunda, 1990; Roeser et al., 2016)。另一方面，道德标准的关注则促使个体对道德与不道德行为进行更加严格和准确的判断，任何不道德行为都更可能反映在其自我概念中(Bering et al., 2005)。此外，道德标准具有自我调节功能，当个体做出不道德行为时，会引发内疚、自责和内心冲突，从而抑制不道德行为的发生(Bandura et al., 1996)。然而，在道德推脱过程中，自我欺骗削弱了道德标准的自我调节功能，缓解了因不道德行为引发的内部冲突，从而助长了更多不道德行为的发生(Batson, 1999; Tenbrunsel & Messick, 2004)。基于此，本研究提出假设 2：道德标准抑制了不道德行为中的自我欺骗的产生。

以往研究通常采用主观报告的量表得分或通过主观报告与行为反应的不一致来测量个体是否产生自我欺骗，但这种主观测量方式可能存在偏差(Chance & Norton, 2015; Sheridan et al., 2015)。随着实验范式的不断发展以及脑科学技术的进步，研究者们开始利用各种脑科学技术来探索欺骗及自我欺骗的内在神经机制，如事件相关电位(ERPs)和功能性磁共振成像(fMRI)。内侧前额叶皮层在自欺发挥重要作用(Abe et al., 2007; Farrow, 2015; Lee et al., 2009)。事件相关电位(Event-related potential, ERP)是一种常用且优秀的技术，用于测量高时间分辨率的结果评估处理下神经反应的时间过程(范伟 等, 2022; Gangl et al., 2017)。因此，本研究通过 ERP 技术探索了不道德行为中道德标准对自欺行为影响的神经反应。基于以往的研究，本研究选择 N1、P2、N2 和 P300 作为检测自我欺骗的潜在生理指标。

首先，N1 成分与决策中的信息加工过程（尤其是对决策刺激的注意过程）有关(Cuthbert et al., 1998)。研究表明，个体投入的注意资源越多，其 N1 波幅就越大(Martin & Potts, 2009)。有研究发现 N1 是对视觉感知刺激反应较为敏感的成分，自我欺骗主要发生在反应的早期阶段，反映了大脑感知觉区域的敏感性(范伟 等, 2022; Jian et al., 2019)。其次，有学者采用 ERP

技术研究反馈对自我欺骗的影响，发现积极反馈和模糊反馈可能促进自我欺骗的产生，且自我欺骗会诱发较大的 P2 成分(范伟 等, 2022; 钟罗金 等, 2019)。P2 与个体的觉醒水平相关，反映了注意力捕捉(Carretie´ et al., 2001; Potts, 2004)。研究还发现，自我欺骗主要激活 P2 成分，这与 P2 波幅反映个体更关注积极、正向结果的观点相一致(范伟 等, 2022; Rottenburger et al., 2019)。因此，P2 成分也可能是衡量自我欺骗的一个指标。再者，研究发现自控力较低的个体更容易表现出更多的欺骗行为，且 N2 波幅更大(Fan et al., 2020)。欺骗通常被认为是一种不道德行为，个体在欺骗时会出现认知和道德的双重冲突，尤其是在个人利益和道德标准之间做出权衡时(Ofen et al., 2016)。因此，欺骗比诚实需要更多的执行控制。关于欺骗行为的脑电研究显示，N2 成分的增大表明在欺骗行为中，个体需要更多的执行控制进而抑制认知和道德的双重冲突(Hu et al., 2015)。N2 成分是认知控制的核心成分，其波幅增大表明大脑在处理冲突信息时更加活跃(Pires et al., 2014)。而自我欺骗可能涉及到复杂的内部冲突，可能表现为 N2 波幅的增大。因此，将 N1、P2 和 N2 成分作为自我欺骗的测量指标是有必要的。此外，与说真话相比，说谎的 P300 振幅减弱(Suchotzki & Crombez, 2015; Wu et al., 2009)。同理，以往研究发现欺骗行为也会诱发更小的 P3 成分(Hu et al., 2015)。大量研究表明，自我控制资源充足比自我控制资源衰竭组诱发了更小的 P3 波幅(Christ et al., 2008; Cui et al., 2017; Fan et al., 2021; Hu et al., 2015; Wu et al., 2009)。以往 ERPs 研究发现，更小的 P3 波幅反映了执行控制的参与，当实验操作增加了执行控制的需求时，P3 波幅将会减小，这些操作包括知觉负荷、双任务、模糊分类和刺激反应不相容等(Debey et al., 2012)。在人类欺骗行为中，认知负荷是识别欺骗行为的重要指标(Von Hippel & Trivers, 2011)。欺骗认知负荷假说认为欺骗具有双重任务特征，需要注意资源(Suchotzki & Crombez, 2015; Vrij & Granhag, 2011)。在欺骗行为中，个体在处理个人利益与道德标准冲突时，需要更多的认知资源来监控和解决冲突。而自我欺骗行为作为人际间欺骗方式存在的依据之一，是具有节省认知资源、减少认知负荷的优点。研究发现，自我欺骗通过损害非随意性意识记忆以减少认知负荷，且高认知负荷环境可能进一步促进自我欺骗的发生(Jian et al., 2019)，这种认知负荷的减少可能与更大的 P300 波幅有关(Yang et al., 2024)。因为自我欺骗的个体不需要在道德和个人利益之间进行复杂的权衡，他们可能会通过自我合理化或忽视道德标准进行快速决策，可能涉及到较少的认知资源投入，表现为 P300 波幅的增大。也有研究发现，P300 也反映了内隐的自我积极偏差(Chen et al., 2014)。因此，本研究将 P300 也纳入衡量自我欺骗的指标中。基于此，本研究提出假设 3：相比于诚实试次，被试在不道德行为中的自我欺骗诱发了更大的 N1、P2、N2 以及 P300。

本研究通过两个实验，运用事件相关电位技术考察不道德行为情境下自我欺骗的神经机制。实验 1 采用发送者-接受者范式，通过分析 N1、P2、N2 以及 P300 成分，探讨自我欺骗认知过程的电生理机制。实验 2 引入道德标准启动任务，探究道德标准对自我欺骗的影响，并通过分析脑电成分的变化，考察道德标准的抑制作用。本研究预期将揭示不道德行为中自我欺骗的脑电活动特征，并阐明道德标准如何调节自我欺骗过程。这将为理解和干预不道德行为提供重要的神经科学依据。

2 实验 1 不道德行为中自我欺骗的内在神经机制

2.1 实验目的与假设

实验 1 考察不道德行为中自我欺骗的内在神经机制。研究假设：（1）欺骗试次中选择小于真实信念的比例显著大于诚实试次中选择小于真实信念的比例。（2）相比于诚实试次，被试在欺骗试次中会诱发更大的 N2 以及 P300 成分。（3）在脑后区，相比较诚实，欺骗试次会诱发更大的 P2 成分。

2.2 研究方法

2.2.1 被试

使用 G-power 3.1 计算所需样本量，在保证效应量 Cohen' s $d = 0.5$ 的前提下，设定$\alpha = 0.05$，至少需要 27 名被试才能达到 80%$(1-\beta)$ 的统计检验力(Faul et al., 2007)。最终招募 30 名湖南师范大学的在校大学生，其中 5 名具有极端数据的被试被剔除（试次中的欺骗比率低于 10%，或者大于 90%），最后对 25 名被试的数据被纳入分析（男 15 名，$M = 21.03 \pm 2.12$ 岁）。所有被试视力或矫正视力正常，并且之前均未参加过类似实验。本实验获得湖南师范大学伦理委员会的认可，并且被试签署实验知情同意书，在实验结束后给予一定的报酬。

2.2.2 实验设计

采用 2（行为决策：欺骗 vs. 诚实）单因素两水平被试内实验设计。因变量为预测小于真实信念的比例以及 ERP 数据的 N1、P2、N2 和 P300 成分。

2.2.3 实验材料

彩票抽奖任务范式：本实验采用彩票抽奖任务范式结合诱发个体主动不道德行为的发送者-接受者任务范式考察不道德行为对自我欺骗影响(Samad, 2021)。选取该任务范式的原因在于：自我欺骗的经典定义认为自我欺骗发生的必要条件是个体在脑海中同时存在两个互

相矛盾的信念，一个真实信念而另一个是虚假信念(Gur & Sackeim, 1979; Pinker, 2011)。彩票抽奖任务范式将被试的真实信念转化为了可以量化和计算的概率值(P_{true})，通过比较被试的真实信念和预测信念可以更加直观的观察到被试是否在发送者-接受者任务的预测中产生了虚假的信念，对于被试是否发生了自我欺骗能做出相对更加直接的推断。

首先，主试通过抽奖任务测试被试对随机概率 P 的真实信念(P_{true}) (Andreoni & Sanchez, 2019; Schotter & Trevino, 2014)（见图 1）。为更加精确地测试被试对随机概率 P 值的真实信念(P_{true})，本研究将选项之间概率的间隔设定为了 5%。在抽奖任务中，屏幕依次呈现一左一右两种彩票进行选择：左侧彩票的获奖概率为随机概率 P（例如，获得 15 元的概率由计算机随机生成），而右侧彩票的获奖概率是明确的（例如，80%获得 15 元，20%获得 0 元）。在抽奖任务中，当被试的选择从右侧彩票切换到左侧彩票时，右侧明确概率的两个相邻值（即切换点的上限值和下限值）的平均值被用作估计被试对随机概率 P 值的真实信念(P_{true})。通过 11 次选择的切换点，我们得到了被试对随机概率 P 值的真实信念(P_{true})。需要说明的是，每位被试的真实信念(P_{true})都是一个具体的数值，且各被试的真实信念互不相同。

其次，在发送者-接受者任务范式中的每个奖金分配方案的试次中，被试根据方案获得自己的奖金，而接受者的奖金则受一个未知随机概率 P 的影响，这意味着接受者有一定概率 P 获得金钱。在每次被试对奖金分配方案进行选择后，被试需要判断接受者获得奖金的随机概率 P 是大于还是小于其自身的真实信念(P_{true})。具体而言，被试按下“F”键表示预测接受者获得奖金的随机概率 P 小于自身的真实信念($P < P_{true}$)，即认为接受者获得金钱的可能性较低。若被试按下“J”键表示预测接受者获得奖金的随机概率 P 大于自身的真实信念($P > P_{true}$)，即认为接受者获得奖金的可能性较高。如果被试在发送者-接受者任务中的行为决策被诱导产生欺骗行为并且在后期预测接受者获得金钱的可能性比较低（即选择 $P < P_{true}$），这种行为可以被解释为一种自我辩解：“我欺骗你并非因为我不道德，而是因为我对随机概率持悲观态度，即使诚实发送对你更有利的分配方案，你也可能拿不到奖金。”的自我辩解。通过记录被试在发送者-接受者任务中对接受者获得奖金的随机概率 P 的判断（大于或小于自身的真实信念 P_{true}），我们可以考察被试是否产生了虚假的信念，从而进一步探讨其是否发生了自我欺骗。

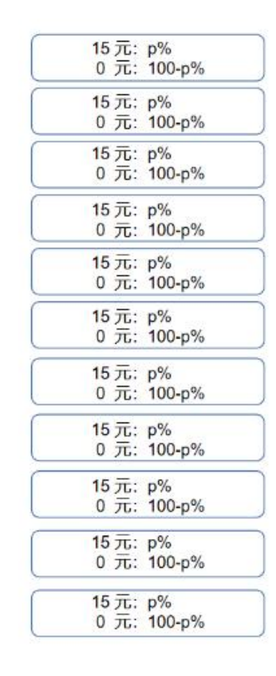


图 1 抽奖任务矩阵图

发送者-接受者任务：发送者-接受者任务范式是近年来广泛应用于研究不诚实和欺骗行为的研究方法之一，具备较高的生态效度(Shuster & Levy, 2020; Zheltyakova & Kireev, 2020)。在该任务中，参与者被分为信息的发送者和接受者两个角色，被试充当发送者。在任务中，发送者和接受者要共同分配一笔奖金，有两个分配方案，一个方案对发送者有利，即发送者能分配到更多的奖金；另一个方案对接受者有利，即接受者能分配到更多的奖金。在实验 1 中的发送者-接受者任务中，主试并未直接要求被试进行欺骗行为，而是诱导被试自主选择是否做出不诚实的欺骗行为，即发生主动不道德行为(Shuster & Levy, 2020; Zheltyakova & Kireev, 2020)。被试被指示要向另一名玩家发送一则信息，信息的内容是“选项____对你而言是更有利的”，并最终按照这一方案分配奖金。接受者只会看到最终发送的方案，看不到具体的分配数额。在任务中，如果被试选择按照要求发送对接受者更有利的方案，则被视为诚实行为；如果被试为了自己得到更多的金钱而选择发送对自己更有利的方案，则被视为欺骗行为。为了让被试在行为决策中有足够的欺骗试次与诚实试次进行分析，实验开始前被试被告知实验过程中发送的选项可以根据自己意愿进行选择，间接提示被试存在欺骗的机会。在实验 1 中，被试总共会进行 180 个 trials 的试次选择（如图 2）。

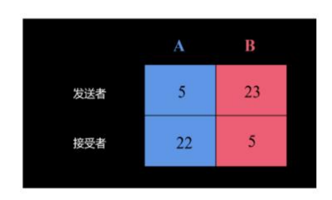




图 2 发送者-接受者任务图

2.2.4 实验程序

整个实验包括两个阶段，被试需要先进行彩票抽奖任务再完成发送者-接受者任务。在被试进行脑电实验的前一天，先完成彩票抽奖任务测试被试对于随机概率的真实信念(P_{true})。主试会告知被试抽奖任务中的奖金是额外的，与第二个阶段任务无关。抽奖任务中，屏幕上会依次呈现一左一右两种彩票供被试选择，其中左边彩票的获奖概率是随机概率 P，右边彩票的获奖概率是明确的，被试一共做 11 次选择（见图 3）。

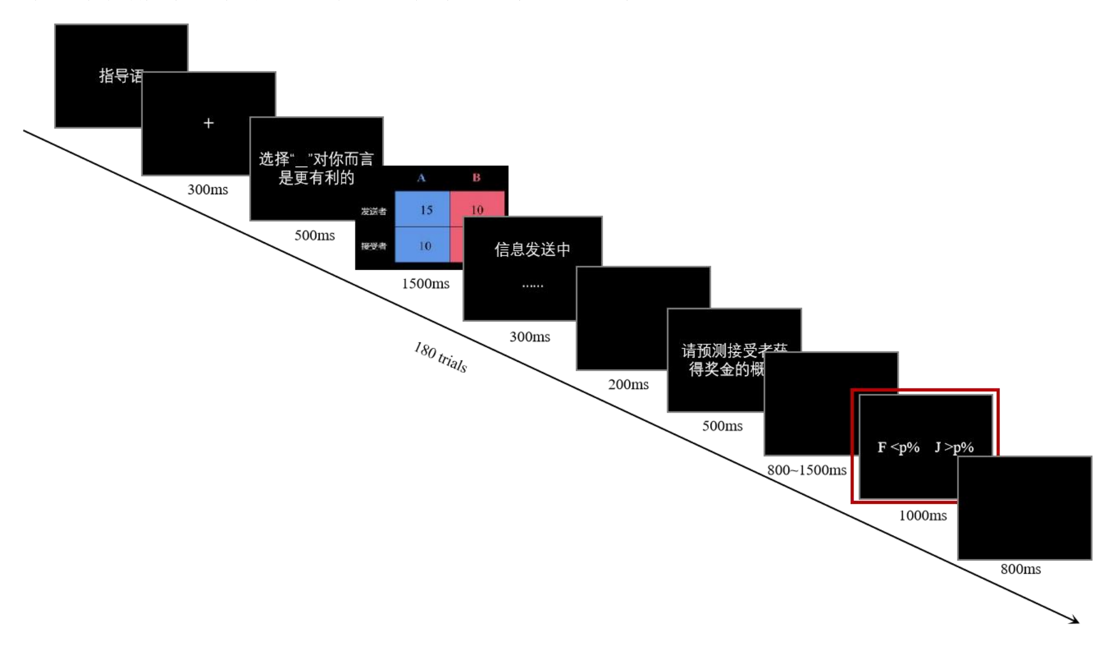


图 3 实验 1 流程图

在实验的第二阶段，被试需要带上电极帽进行发送者-接受者任务并收集脑电数据。主试首先告知被试，他们需要完成一个双人互动的信息发送任务，并安排实验助手扮演另一名玩家与被试见面。接着，主试将两人分开，把被试带到一个单独的房间，并让被试坐在电脑桌前准备开始任务。主试随后讲述任务规则：本次任务中有两名角色，一名是信息的发送者（被试本人），另一名是信息的接受者（实验助手）。每次任务中，双方会共同分配一笔奖金，提供两个分配方案选项 A 和 B 供被试选择。被试需要向接受者发送一则信息：“选项____对你而言是更有利的”，并选择 A 或者 B 方案发送给接受者。无论被试选择哪个方案发送，最终会按照所选方案进行奖金分配，接受者只会看到最终发送的分配方案。在奖金分配时，被试会获得应得的奖金，而接受者获得应得奖金的概率是(P)，反之有(1-P)的概率什么也得不到。被试在选择完发送信息的方案之后，主试会要求被试对接受者获得奖金的概率进行判断，包括两个选项：$P<P_{true}$ 和 $P>P_{true}$，且决策不会影响双方的任何收益。被试经过练习阶段以熟悉任务规则后正式开始实验，在每次试次任务中所累积的奖金最终会按固定比例换算成人民币作为任务的奖励。实验开始前有 10 个练习试次，正式实验包含 180 个试次。初始时，屏幕中呈现 300ms 的注视点，然后呈现 500ms 的信息界面，显示即将发送给接受者的信息。接着呈现 A 和 B 两种分配方案供被试选择，选择 A 方案按“F”键，选择 B 方

案按“J”键，呈现时间为1500ms。若被试按“F”键，则意味着发送给接受者的信息为“选项A对你而言是更有利的”。若被试按“J”键，则意味着发送给接受者的信息为“选项J对你而言是更有利的”。被试完成选择后进入300ms的等待界面，显示信息发送中。如果被试在规定时间内未进行方案选择，则显示系统正随机发送信息，200ms 的黑屏后进入预测界面。在脑电实验的预测阶段，将前一天测试被试的真实信念(P_{true})作为参照，被试会在接受者获得奖金的随机概率(P)是大于还是小于自身的真实信念(P_{true})。若被试按“F”键，则意味着被试认为接受者获得奖金的随机概率(P)小于自身的真实信念(P_{true})。若被试按“J”键，则意味着被试认为接受者获得奖金的随机概率(P)大于自身的真实信念(P_{true})，做出选择后继续进入下一个试次。

2.2.5 数据收集与分析

（1）行为数据

实验采用E-prime 2.0呈现实验程序并完成数据采集。数据预处理中，为了确保统计效应，被试在所有试次中选择诚实和欺骗的试次不能过少(Shuster & Levy, 2020; Zheltyakova & Kireev, 2020)。依据前人文献，本实验纳入欺骗率在10%~90%之间的被试数据，剔除欺骗率小于10%或大于90%的数据。数据处理和呈现中，将被试在抽奖任务中从右边选项切换到左边选项时的两个概率的平均数记为真实信念(P_{true})。将被试在发送者-接受者任务范式中接受者获得的奖金记为未知随机概率(P)。数据统计分析中，本实验对欺骗试次和诚实试次的选择“$P<P_{true}$”选项的比例进行配对样本 t 检验。

（2）脑电数据记录和分析

本研究采用64通道的放大器(ANT Neuro, Enschede, Netherlands)和64导电极帽进行脑电记录，电极排列符合国际10-20系统定位标准。脑电信号记录时以CP_Z点作为参考电极，双侧乳突M1以及M2作为离线参考。所有电极与头皮之间的阻抗都小于5 kΩ，采样率为500 Hz/导。使用 MATLAB 2017b 采集脑电图数据和EEGLAB14.1.2工具箱(Delorme & Makeig, 2004; Plöchl et al., 2012)进行数据分析。首先，对数据进行滤波，滤波工具是EEGLAB 工具包内置的 Hamming windowed sinc FIR(finite impulse response)滤波器，参数为 0.1~30 Hz (filter slopes: 24 dB/octave)。然后，利用 ICA(独立成分分析, Independent Component Analysis)方法去除脑电中的水平和垂直眼电和伪迹(Delorme & Makeig, 2004)。波幅大于$\pm 70\mu V$视为伪迹而自动剔除。之后，对数据进行分段，从每个决策界面呈现前200 ms至800 ms的连续数据文件中提取epoch（见图3中的“做出选择”）。决策界面呈现前−200 ms至0 ms时间窗口内的活动作为每个ERP的基线。

在本研究中，如果参与者选择了“$P<P_{true}$”选项，它被认为是分析自我欺骗发生的有效

试次。根据本研究目的、脑地形图及视觉检测，我们分析的是时间窗 N1（20~80ms）、P2（100~200ms）、N2（150~250ms）和 P300（200~300ms）的平均成分。本研究进行两因素被试内方差分析 2（行为决策：诚实 vs. 欺骗）×5[脑区：额区(F3, Fz, F4) vs. 额中区(FC3, FCz, FC4) vs. 中央区(C3, Cz, C4) vs. 中顶区(CP3, CPz, CP4) vs. 顶区(P3, Pz, P4)]；以及 2（行为决策：诚实 vs. 欺骗）×3[脑半球：左半球(F3、FC3、C3、CP3、P3) vs. 中央区(Fz、FCz、Cz、CPz、Pz) vs. 右半球(F4、FC4、C4、CP4、P4)]。主效应和交互效应的 p 值采用 Greenhouse-Geisser 方法对违反球度假设的 p 值进行校正，对多重比较采用 Bonferroni 校正。

2.3 研究结果

2.3.1 行为结果

采用配对样本 t 检验，对被试在欺骗试次中预测小于真实信念的比例($P_d < P_{true}$)与被试在诚实试次中预测小于真实信念的比例($P_h < P_{true}$)进行显著性检验。结果发现，欺骗试次中的预测信念小于真实信念的比例($P_d < P_{true}$) ($M \pm SD$ = 62.12% ± 15.36%)显著大于诚实试次中的预测信念小于真实信念的比例($P_h < P_{true}$) ($M \pm SD$ = 51.12% ± 12.74%)，$t(24) = 3.09$，$p = 0.005$，Cohen's $d = 0.78$，95%CI = [3.66, 18.35] （见图 4）。

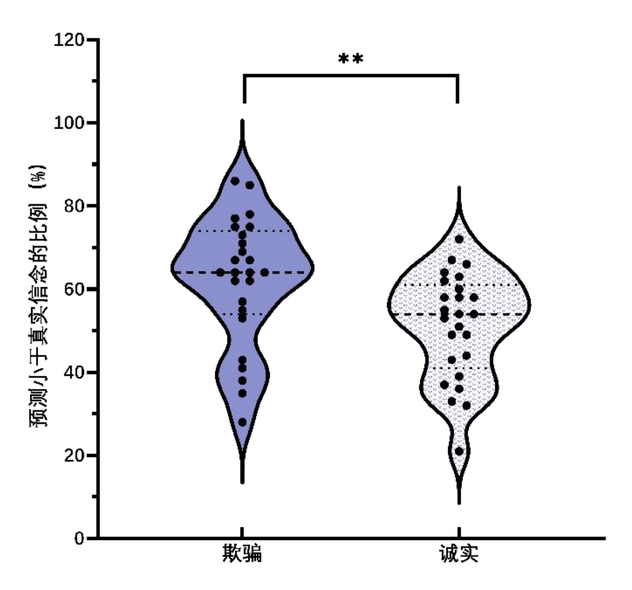


图 4 不同行为决策条件下预测小于真实信念的比例

2.3.2 ERP 结果

（1）N1(20~80ms)

通过对 N1 成分进行 2（行为决策：诚实 vs. 欺骗）×5（脑区：额区 vs. 额中区 vs. 中央区 vs. 中顶区 vs. 顶区）的重复测量方差分析可知，行为决策的主效应不显著，$F(1, 24) = 0.01$，$p = 0.924$。脑区的主效应具有显著的差异，$F(4, 96) = 11.84$，$p < 0.001$，$\eta_p^2 = 0.33$。

通过事后多重比较发现，额区($M \pm Standard\ Error\ [SE] = 0.87 \pm 0.23 \mu V$)诱发的 N1 成分显著大于中顶区($M \pm SE = 0.25 \pm 0.17 \mu V$)，$p = 0.007$，95%$CI$ = [0.12, 1.11]; 和顶区($M \pm SE = 0.12 \pm 0.18 \mu V$)，$p = 0.015$，95%$CI$ = [0.10, 1.40]。额中区($M \pm SE = 0.70 \pm 0.20 \mu V$)诱发的 N1 成分显著大于中顶区($M \pm SE = 0.25 \pm 0.17 \mu V$)，$p = 0.006$，95%$CI$ = [0.10, 0.80]; 和顶区($M \pm SE = 0.12 \pm 0.18 \mu V$)，$p = 0.027$，95%$CI$ = [0.04, 1.12]。中央区($M \pm SE = 0.53 \pm 0.18 \mu V$)诱发的 N1 成分显著大于中顶区($M \pm SE = 0.25 \pm 0.17 \mu V$)，$p = 0.001$，95%$CI$ = [0.09, 0.47]; 和顶区($M \pm SE = 0.12 \pm 0.18 \mu V$)，$p = 0.034$，95%$CI$ = [0.02, 0.81]。行为决策与脑区的交互作用不显著，$F(4, 96) = 3.33$，$p = 0.061$（见图 5）。

通过对 N1 成分进行 2（行为决策：诚实 vs. 欺骗）×3（脑半球：左半球 vs. 中央区 vs. 右半球）的重复测量方差分析可知，行为决策的主效应不显著，$F(1, 24) = 0.009$，$p = 0.924$。脑半球的主效应具有显著性的差异，$F(2, 48) = 15.35$，$p < 0.001$，$\eta_p^2 = 0.390$。通过事后比较发现，左半球($M \pm SE = 0.70 \pm 0.19 \mu V$)诱发的 N1 成分显著大于右半球($M \pm SE = 0.21 \pm 0.17 \mu V$)，$p = 0.001$，95%$CI$ = [0.18, 0.80]；中央区($M \pm SE = 0.57 \pm 0.18 \mu V$)诱发的 N1 成分显著大于右半球($M \pm SE = 0.21 \pm 0.17 \mu V$)，$p < 0.001$，95%$CI$ = [0.19, 0.54]。行为决策与脑半球的交互作用不显著，$F(2, 48) = 0.73$，$p = 0.485$（见图 5）。

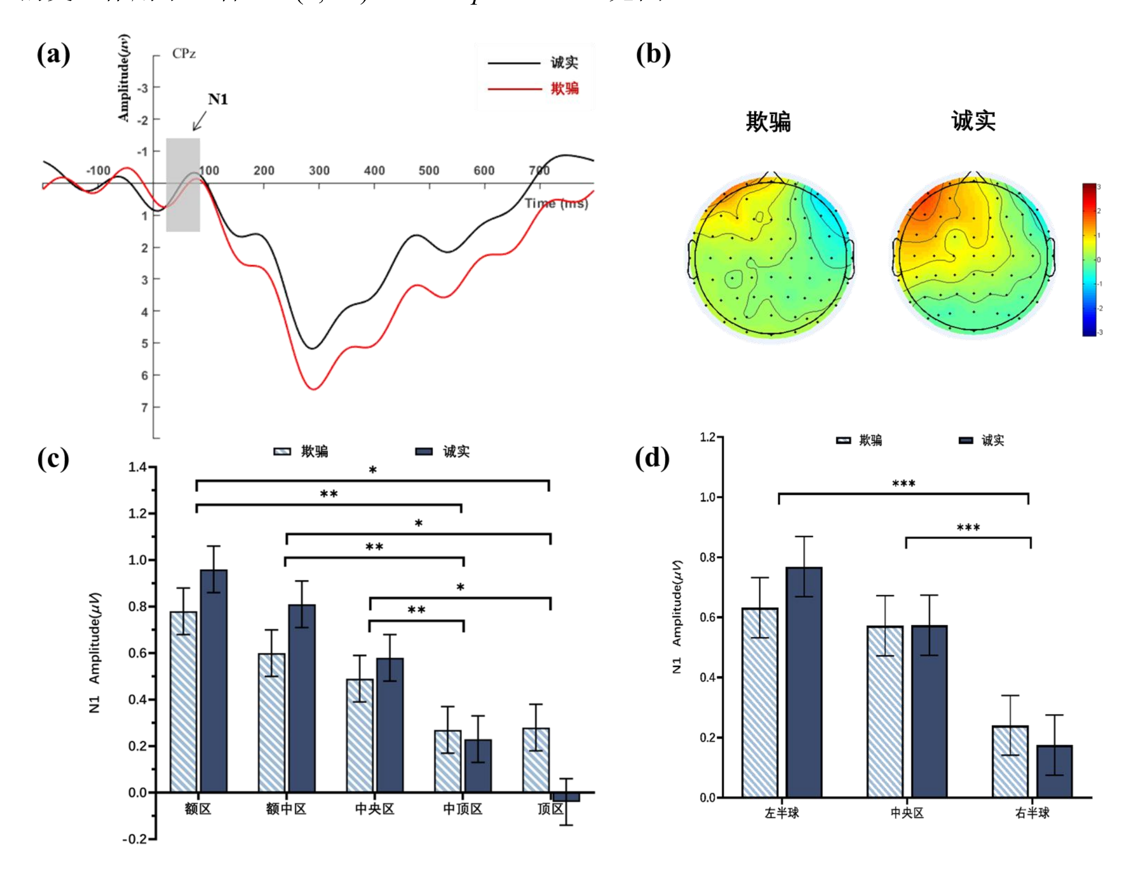


图 5 (a) CPz 在不同行为决策条件下的平均 ERPs，图中灰色条表示 N1 的时间窗(20~80ms)。(b)每种条件下 N1 的脑地形图。(c、d)柱状图显示了不同脑区和半球下欺骗和诚实条件下的平均 N1 值，误差条表示平均值的标准误差。

（2）P2(100~200ms)

通过对 P2 成分进行 2（行为决策：诚实 vs. 欺骗）×5（脑区：额区 vs. 额中区 vs. 中央区 vs. 中顶区 vs. 顶区）的重复测量方差分析可知，行为决策的主效应不显著，$F(1, 24) = 3.51$，$p = 0.073$。脑区的主效应不显著，$F(4, 96) = 0.96$，$p = 0.431$。行为决策与脑区存在交互作用，$F(4, 96) = 5.23$，$p = 0.001$，$\eta_p^2 = 0.18$。简单效应分析表明，在中顶区，相比较诚实的条件($M \pm SE$= 1.30 ± 0.28μV)，欺骗条件($M \pm SE$= 2.18 ± 0.45μV)诱发了更大的 P2 成分，$p = 0.043$，95%CI = [−1.74, −0.03]。在顶区，相比较诚实的条件($M \pm SE$= 0.84 ± 0.33μV)，欺骗条件($M \pm SE$= 2.14 ± 0.57μV)诱发了更大的 P2 成分，$p = 0.002$，95%CI = [−2.07, −0.52]（见图 6）。

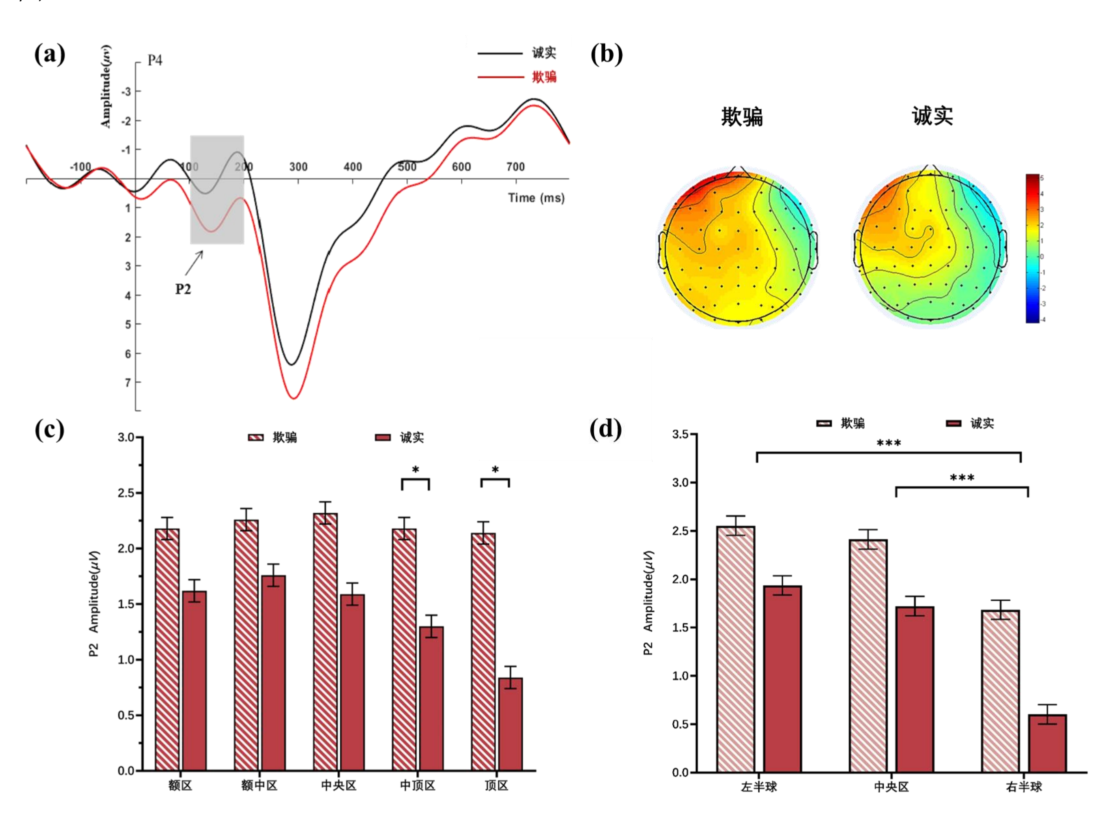


图 6 (a) P4 在不同行为决策条件下的平均 ERPs，图中灰色条表示 P2 的时间窗(100~200ms)。(b)每种条件下 P2 的脑地形图。(c、d)柱状图显示了不同脑区和半球下欺骗和诚实条件下的平均 P2 值，误差条表示平均值的标准误差。

通过对 P2 成分进行 2（行为决策：诚实 vs. 欺骗）×3（脑半球：左半球 vs. 中央区 vs. 右半球）的重复测量方差分析可知，行为决策的主效应不显著，$F(1, 24) = 3.51$，$p = 0.073$。

脑半球的主效应具有显著性的差异，$F(2, 48) = 27.82$，$p < 0.001$，$\eta_p^2 = 0.54$。通过事后比较发现，左半球($M \pm SE = 2.25 \pm 0.29\mu V$)诱发的P2成分显著大于右半球($M \pm SE = 1.14 \pm 0.27\mu V$)，$p < 0.001$，95%$CI$ = [0.67, 1.53]；中央区($M \pm SE = 2.07 \pm 0.30\mu V$)诱发的 P2 成分显著大于右半球($M \pm SE = 1.14 \pm 0.27\mu V$)，$p < 0.001$，95%$CI$ = [0.57, 1.28]。行为决策与脑半球的交互作用不显著，$F(2, 48) = 2.48$，$p = 0.094$（见图 6）。

（3）N2(150~250ms)

通过对 N2 成分进行 2（行为决策：诚实 vs. 欺骗）×5（脑区：额区 vs. 额中区 vs. 中央区 vs. 中顶区 vs. 顶区）的重复测量方差分析可知，行为决策的主效应显著，$F(1, 24) = 6.56$，$p = 0.017$，$\eta_p^2 = 0.22$。通过事后比较发现，欺骗试次($M \pm SE = 3.26 \pm 0.43\mu V$)诱发的 N2 成分显著大于诚实试次($M \pm SE = 2.03 \pm 0.40\mu V$)，$p = 0.017$，95%$CI$ = [0.24, 2.21]。脑区的主效应具有显著的差异，$F(4, 96) = 8.84$，$p < 0.001$，$\eta_p^2 = 0.27$。通过事后比较发现，额中区($M \pm SE = 3.30 \pm 0.45\mu V$)诱发的 N2 成分显著大于中顶区($M \pm SE = 2.23 \pm 0.31\mu V$)，$p = 0.016$，95%$CI$ = [0.14, 2.00]。额中区($M \pm SE = 3.30 \pm 0.45\mu V$)诱发的 N2 成分显著大于顶区($M \pm SE = 1.57 \pm 0.34\mu V$)，$p = 0.025$，95%$CI$ = [0.15, 3.32]。中央区($M \pm SE = 2.96 \pm 0.37\mu V$)诱发的 N2 成分显著大于中顶区($M \pm SE = 2.23 \pm 0.31\mu V$)，$p = 0.002$，95%$CI$ = [0.21, 1.25]。中央区($M \pm SE = 2.96 \pm 0.37\mu V$)诱发的 N2 成分显著大于顶区($M \pm SE = 1.57 \pm 0.34\mu V$)，$p = 0.012$，95%$CI$ = [0.23, 2.55]。行为决策与脑区的交互作用不显著，$F(4, 96) = 0.33$，$p = 0.856$（见图 7）。

通过对 N2 成分进行 2（行为决策：诚实 vs. 欺骗）×3（脑半球：左半球 vs. 中央区 vs. 右半球）的重复测量方差分析可知，行为决策的主效应显著，$F(1, 24) = 6.56$，$p = 0.017$，$\eta_p^2 = 0.22$。通过事后比较发现，欺骗试次($M \pm SE = 3.26 \pm 0.43\mu V$)诱发的 N2 成分显著大于诚实试次($M \pm SE = 2.03 \pm 0.40\mu V$)，$p = 0.017$，95%$CI$ = [0.24, 2.21]。脑半球的主效应具有显著性的差异，$F(2, 48) = 10.98$，$p < 0.001$，$\eta_p^2 = 0.31$。通过事后比较发现，左半球($M \pm SE = 2.88 \pm 0.36\mu V$)诱发的 N2 成分显著大于右半球($M \pm SE = 2.00 \pm 0.35\mu V$)，$p = 0.027$，95%$CI$ = [0.08, 1.67]。中央区($M \pm SE = 3.05 \pm 0.39\mu V$)诱发的 N2 成分显著大于右半球($M \pm SE = 2.00 \pm 0.35\mu V$)，$p < 0.001$，95%$CI$ = [0.52, 1.59]。行为决策与脑半球的交互作用不显著，$F(2, 48) = 0.10$，$p = 0.910$（见图 7）。

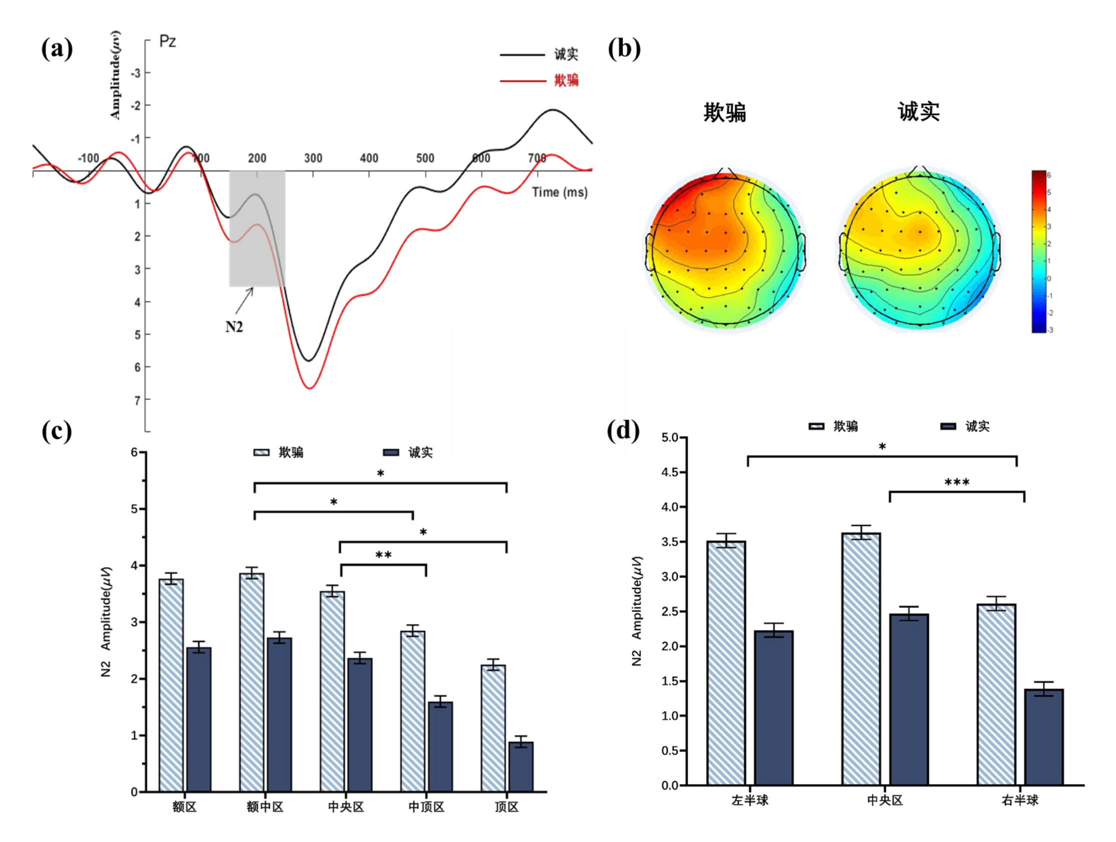


图 7 (a) Pz 在不同行为决策条件下的平均 ERPs，图中灰色条表示 N2 的时间窗(150~250ms)。(b)每种条件下 N2 的脑地形图。(c、d)柱状图显示了不同脑区和半球下欺骗和诚实条件下的平均 N2 值，误差条表示平均值的标准误差。

（4）P300(200~300ms)

通过对 P300 成分进行 2（行为决策：诚实 vs. 欺骗）×5（脑区：额区 vs. 额中区 vs. 中央区 vs. 中顶区 vs. 顶区）的重复测量方差分析可知，行为决策的主效应显著，$F(1, 24) = 6.99$，$p = 0.014$，$\eta_p^2 = 0.23$。通过事后比较发现，欺骗试次($M \pm SE = 5.08 \pm 0.47\mu V$)诱发的 P300 成分显著大于诚实试次($M \pm SE = 3.67 \pm 0.46\mu V$)，$p = 0.014$，95%$CI$ = [0.31, 2.50]。脑区的主效应不显著，$F(4, 96) = 2.43$，$p = 0.053$。行为决策与脑区的交互作用不显著，$F(4, 96) = 2.15$，$p = 0.081$（见图 8）。

通过对 P300 成分进行 2（行为决策：诚实 vs. 欺骗）×3（脑半球：左半球 vs. 中央区 vs. 右半球）的重复测量方差分析可知，行为决策的主效应具有显著性的差异，$F(1, 24) = 6.99$，$p = 0.014$，$\eta_p^2 = 0.23$。通过事后比较发现，欺骗试次($M \pm SE = 5.08 \pm 0.47\mu V$)诱发的 P300 成分显著大于诚实试次($M \pm SE = 3.67 \pm 0.46\mu V$)，$p = 0.014$，95%$CI$ = [0.31, 2.50]。脑半球的主效应不具有显著性的差异，$F(2, 48) = 1.10$，$p = 0.341$。行为决策与脑半球的交互作用不显著，$F(2, 48) = 0.23$，$p = 0.798$（见图 8）。

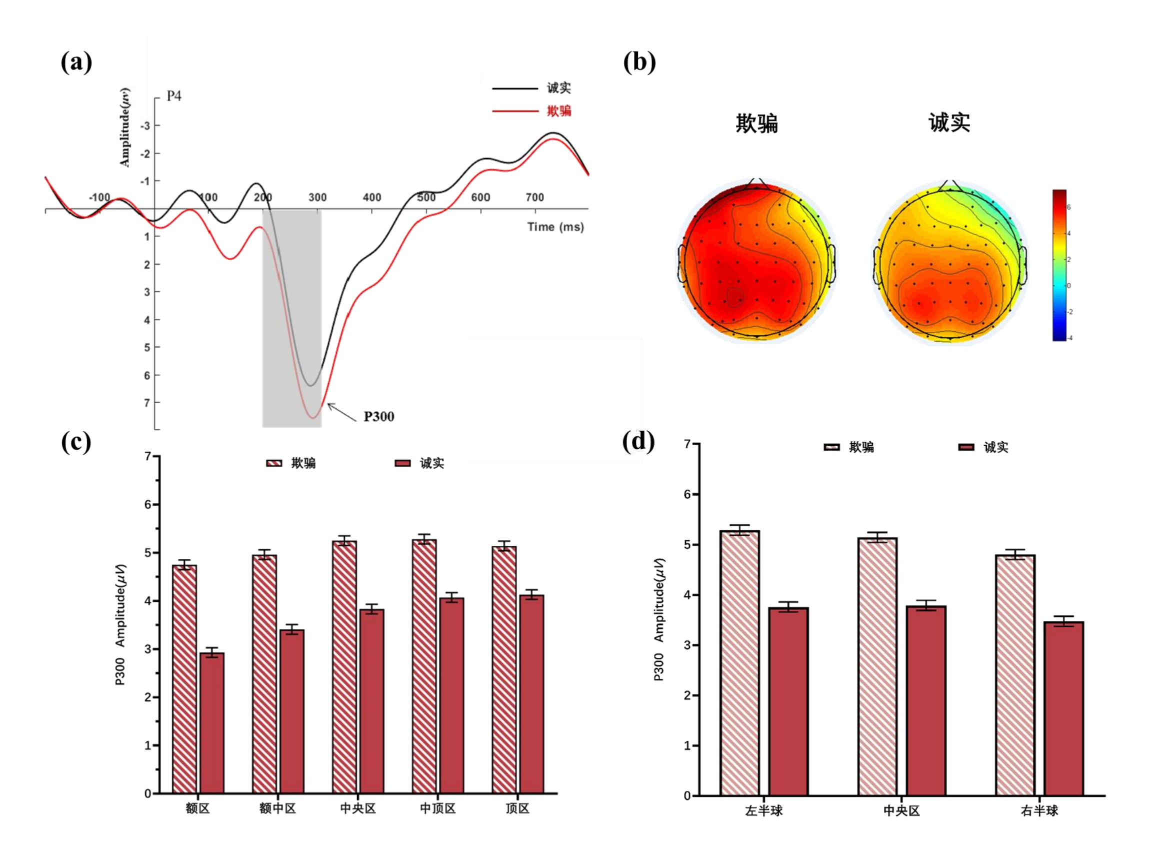


图 8 (a) P4 在不同行为决策条件下的平均 ERPs，图中灰色条表示 P300 的时间窗(200~300ms)。(b)每种条件下 P300 的脑地形图。(c、d)柱状图显示了不同脑区和半球下欺骗和诚实条件下的平均 P300 值，误差条表示平均值的标准误差。

2.4 实验 1 讨论

实验 1 的行为结果表明，在欺骗试次中，被试选择低于真实信念的比例显著高于在诚实试次中的比例。这些结果可能表明个体在不道德行为中倾向于采取比真实信念更小的虚假信念来维护道德自我，通过对随机概率值的悲观解释，个体合理化自己的不道德行为，即“我欺骗你不是我不道德，而是我觉得你得到奖金的概率很低，不如让我获得更多的奖金”，这些结果进一步扩宽了道德褪色理论的使用范围(Moore, 2016; Tenbrunsel & Messick, 2004)，验证了假设 1。

ERPs 结果显示，不管欺骗试次还是诚实试次，个体在自我欺骗决策过程中都出现了 N1 和 P2 成分。正如前言部分所述，N1 成分是出现在额叶区域的负向电位成分，其波幅的增大受注意力的影响较大(赵仑, 2010)。研究结果揭示了个体在额区、额中区和中央区诱发更大的 N1 成分，这表明在决策早期阶段，个体对刺激的资源投入较多(Martin & Potts, 2009)。在决策早期阶段，个体可能会经历内心的冲突，这种冲突在 N1 成分的变化中得到了体现。前人研究发现冲突情境比互利情境诱发更大的 N1 波幅(Boudreau et al., 2009)。同时，左侧大脑的 N1 成分、P2 成分和 N2 成分激活更为明显，这可能表明在决策过程中，左侧大脑区域在

注意力分配上扮演着更为重要的角色(Carretie´ et al., 2001; 王益文, 林崇德, 2005)。进一步的研究表明，在中顶区和顶区，欺骗试次相较于诚实试次的自我欺骗诱发了更大的 P2 成分，验证了假设 3。这一发现表明，在欺骗试次中，处于自我欺骗个体更关注积极、正向结果的观点(范伟 等, 2022; Rottenburger et al., 2019)。在欺骗试次中，处于自我欺骗的个体在决策界面时的大脑对于视觉刺激的处理投入了更多的注意资源。P2 成分作为一种早期的判断成分，涉及到视觉刺激的早期加工过程，被认为是注意力有效性的客观指标之一(Carretie´ et al., 2001)。同时 P2 成分也反映了大脑对情绪的认知加工过程。在自我欺骗的情况下，个体需要更多的情绪性动机参与(范伟 等, 2022)。ERPs 结果显示，相比于诚实试次，在欺骗试次中处于自我欺骗的个体会诱发更大的 N2，验证了假设 3。这可能反映了自我欺骗的个体在意识层面也面临着处理个人利益与道德标准的冲突，需要更多的认知资源来监控和解决冲突(Ofen et al., 2016)。在不道德行为中，自我欺骗被当做一种策略来处理个人利益和道德标准的冲突，需要投入更多的认知资源在冲突的监控以及解决过程中(Moore, 2016; Tenbrunsel & Messick, 2004)。此外，相比于诚实试次，被试在欺骗试次中的自我欺骗会诱发更大的 P300 波幅，验证了假设 3。自我欺骗不仅涉及执行控制功能，还反映了复杂的高级认知过程，如决策和记忆(范伟 等, 2022; Hu et al., 2015)。自我欺骗可能通过减少人们的认知负荷而更好地进行欺骗，可能会诱发更大的 P300 波幅，这与以往的研究保持了一致(范伟 等, 2017; Yang et al., 2024)。

3 实验 2 道德标准对不道德行为中自我欺骗神经机制的影响

3.1 实验目的与假设

实验 2 采用 ERP 技术考察关注道德标准影响不道德行为中自我欺骗的内在神经机制。研究假设：（1）在道德标准启动条件下，被试在欺骗试次中选择“$P<P_{true}$”比例与诚实试次中选择“$P<P_{true}$”比例差异不显著。在控制条件下，被试在欺骗试次中选择“$P<P_{true}$”比例显著大于诚实试次中选择“$P<P_{true}$”比例。（2）在道德标准启动条件下，欺骗试次诱发的 P2 显著小于诚实试次。在控制条件下，欺骗试次诱发的 P2 成分与诚实试次无显著差异。在道德标准启动条件下，欺骗试次诱发的 N2 成分与诚实试次无显著差异。在控制条件下，欺骗试次诱发的 N2 显著小于诚实试次。

3.2 研究方法

3.2.1 被试

使用 G-power 3.1 计算所需样本量，在保证效应量 $f = 0.3$ 的前提下，设定 $\alpha = 0.05$，至少需要 24 名被试才能达到 80%$(1-\beta)$ 的统计检验力(Faul et al., 2007)。最终招募 30 名湖南师范大学的在校大学生，其中 5 名具有极端数据的被试被剔除（试次中的欺骗比率低于 10%，或者大于 90%）。最后对 25 名被试的数据被纳入分析（男 15 名，$M = 21.53 \pm 2.34$ 岁）。所有被试视力或矫正视力正常，并且之前均未参加过类似实验。本实验获得湖南师范大学伦理委员会的认可，并且被试签署实验知情同意书，在实验结束后给予一定的报酬。

3.2.2 实验设计

采用 2（组别：关注道德标准组 vs. 控制组）× 2（行为决策：欺骗 vs. 诚实）两因素被试内实验设计。因变量为预测小于真实信念的比例以及 ERP 数据的 N1、P2、N2 和 P300 成分。

3.2.3 实验材料

彩票抽奖任务范式：同实验 1。

发送者-接受者任务：同实验 1。

关注道德标准启动材料：前人的研究表明，通过回忆任务可以成功地启动对道德标准的关注(Schotter & Trevino, 2014)。在该任务中，关注道德标准组的被试被要求在 5 分钟内写下法律条文中禁止做的十件事情（包含道德提醒）。而控制组的被试被要求在同样的时间内写下读过的十本书的名字（没有道德提醒）。

3.2.4 实验程序

实验流程同实验 1。整个实验包括两个阶段，被试需要先进行彩票抽奖任务再完成发送者-接受者任务。在被试进行脑电实验的前一天，先完成彩票抽奖任务测试被试对于随机概率的真实信念(P_{true})。

实验第二阶段任务分为两个 Blocks（关注道德标准条件与控制条件）。被试需要带上电极帽进行发送者-接受者任务并收集脑电数据。在 *I*-Block 中，主试要求被试在 5 分钟内写下读过的十本书的名字（没有道德提醒）。被试完成后，将进行道德标准启动的操作检验，回答一个问题："在刚才的任务中，你认为你对道德标准的关注程度是？"，并进行 7 点评分（1 代表极低，7 代表极高）。随后，被试开始发送者-接受者的任务。在 *II*-Block 中，主试要求被试在 5 分钟内写下法律条文中禁止做的十件事情（包含道德提醒）。同样进行道德标准启动的操作检验并开始发送者-接受者任务。实验开始前有 10 个练习试次。整个任务包括 360 个试验（*I*-Block 包括 180 个试验，*II*-Block 包括 180 个试验，见图 9）。然后开始

正式实验阶段，被试需要在 A 和 B 两个方案中选择一个方案进行信息的发送，然后进入预测界面，与实验 1 一致。将提前测试的被试的真实信念(P_{true})作为参照，主试会要求被试对接受者获得奖金的概率进行判断，包括两个选项：$P<P_{true}$ 和 $P>P_{true}$。实验的流程如图 9 所示。

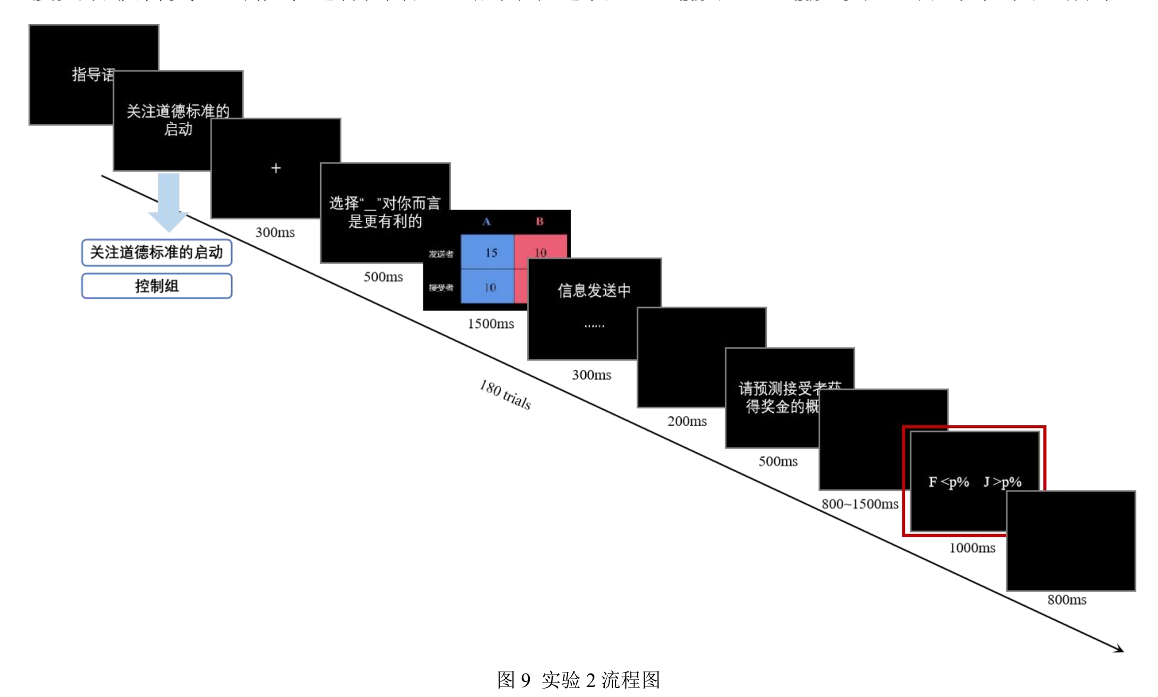


图 9 实验 2 流程图

3.2.5 数据收集与分析

（1）行为数据

实验采用 E-prime 2.0 呈现实验程序并完成数据采集。数据预处理中，为了确保统计效应，被试在所有试次中选择诚实和欺骗的试次不能过少(Shuster & Levy, 2020; Zheltyakova & Kireev, 2020)。依据前人文献，本实验纳入欺骗率在 10%~90%之间的被试数据，剔除欺骗率小于 10%或大于 90%的数据。数据处理和呈现中，将被试在抽奖任务中从右边选项切换到左边选项时的两个概率的平均数记为真实信念(P_{true})。将被试在发送者-接受者任务范式中接受者获得的奖金记为未知随机概率(P)。数据统计分析中，本实验对欺骗试次和诚实试次的选择“$P<P_{true}$”选项的比例进行 2（组别：关注道德标准组 vs. 控制组）× 2（行为决策：欺骗 vs. 诚实）的重复测量方差分析。

（2）脑电数据记录和分析

脑电数据处理的过程与实验 1 一致。在本研究中，如果参与者选择了“$P<P_{true}$”选项，它被认为是分析自我欺骗发生的有效试次。根据本研究目的、脑地形图及视觉检测，我们分析的是时间窗 N1（20~80ms）、P2（100~200ms）、N2（150~250ms）和 P300（200~300ms）的平均成分。本研究进行三因素被试内方差分析 2（组别：关注道德标准组 vs. 控制组）× 2（行为决策：诚实 vs. 欺骗）×5[脑区：额区(F3, Fz, F4) vs. 额中区(FC3, FCz, FC4) vs. 中

央区(C3, Cz, C4) vs. 中顶区(CP3, CPz, CP4) vs. 顶区(P3, Pz, P4)]；以及 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：诚实 vs. 欺骗）×3[脑半球：左半球(F3、FC3、C3、CP3、P3) vs. 中央区(Fz、FCz、Cz、CPz、Pz) vs. 右半球(F4、FC4、C4、CP4、P4)]。主效应和交互效应的 p 值采用 Greenhouse-Geisser 方法对违反球度假设的 p 值进行校正，对多重比较采用 Bonferroni 校正。

3.3 研究结果

3.3.1 行为结果

采用 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：诚实 vs. 欺骗）对预测小于真实信念的比例($P< P_{true}$)进行重复测量方差分析，结果发现，组别的主效应不显著，$F(1, 24) = 1.81$，$p = 0.191$。行为决策的主效应不显著，$F(1, 24) = 1.66$，$p = 0.210$。组别与行为决策的交互作用显著，$F(1, 24) = 6.34$，$p = 0.019$，$\eta_p^2 = 0.21$。简单效应分析表明，在关注道德标准组，欺骗试次中的预测信念小于真实信念的比例($P_d < P_{true}$) ($M \pm SE = 53.92\% \pm 2.60\%$)与诚实试次中的预测信念小于真实信念的比例($P_h < P_{true}$)无显著差异 ($M \pm SE = 53.32\% \pm 2.64\%$)，$p = 0.853$。在控制组，欺骗试次中的预测信念小于真实信念的比例($P_d < P_{true}$) ($M \pm SE = 61.40\% \pm 2.73\%$)显著大于诚实试次中的预测信念小于真实信念的比例($P_h < P_{true}$)($M \pm SE = 52.84\% \pm 2.84\%$)，$p = 0.039$，95%$CI = [0.46, 16.66]$ （见图 10）。

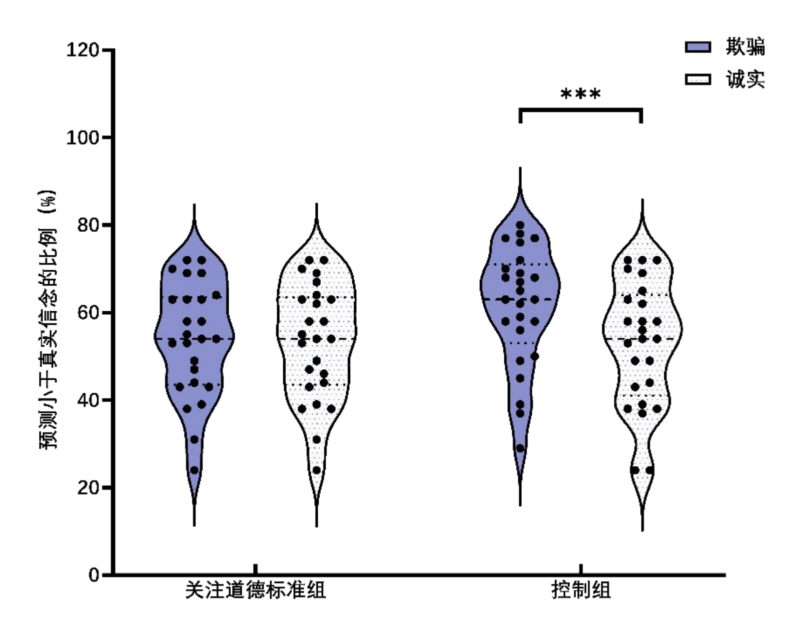


图 10 不同条件下预测小于真实信念的比例

3.3.2 ERP 结果

（1）N1(20~80ms)

通过对 N1 成分进行 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：欺骗 vs. 诚实）×5（脑区：额区 vs. 额中区 vs. 中央区 vs. 中顶区 vs. 顶区）的重复测量方差分析可知，组别的主效应不显著，$F(1, 24) = 1.08$，$p = 0.310$。行为决策的主效应不显著，$F(1, 24) = 0.51$，$p = 0.482$。脑区的主效应具有显著的差异，$F(4, 96) = 18.09$，$p < 0.001$，$\eta_p^2 = 0.43$。通过事后多重比较发现，额区($M \pm SE = 0.70 \pm 0.19\mu V$)诱发的 N1 成分显著大于中央区($M \pm SE = 0.34 \pm 0.16\mu V$)，$p = 0.005$，95%$CI$ = [0.08, 0.63]; 中顶区($M \pm SE = 0.12 \pm 0.15\mu V$)，$p < 0.001$，95%$CI$ = [0.24, 0.93]; 和顶区($M \pm SE = -0.01 \pm 0.14\mu V$)，$p = 0.001$，95%$CI$ = [0.23, 1.18]。额中区($M \pm SE = 0.52 \pm 0.18\mu V$)诱发的 N1 成分显著大于中央区($M \pm SE = 0.34 \pm 0.16\mu V$)，$p = 0.014$，95%$CI$ = [0.03, 0.33]; 中顶区($M \pm SE = 0.12 \pm 0.15\mu V$)，$p = 0.001$，95%$CI$ = [0.15, 0.67]; 和顶区($M \pm SE = -0.01 \pm 0.14\mu V$)，$p = 0.005$，95%$CI$ = [0.12, 0.94]。中央区($M \pm SE = 0.34 \pm 0.16\mu V$)诱发的 N1 成分显著大于中顶区($M \pm SE = 0.12 \pm 0.15\mu V$)，$p < 0.001$，95%$CI$ = [0.09, 0.36]; 和顶区($M \pm SE = -0.01 \pm 0.14\mu V$)，$p = 0.016$，95%$CI$ = [0.05, 0.65]。其他的条件下, 均为无显著差异，$p > 0.05$（见图 11）。

通过对 N1 成分进行 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：欺骗 vs. 诚实）×3（脑半球：左半球 vs. 中央区 vs. 右半球）的重复测量方差分析可知，组别的主效应不显著，$F(1, 24) = 1.08$，$p = 0.310$。行为决策的主效应不显著，$F(1, 24) = 0.51$，$p = 0.482$。脑半球的主效应具有显著性的差异，$F(2, 48) = 20.46$，$p < 0.001$，$\eta_p^2 = 0.46$。通过事后比较发现，左半球($M \pm SE = 0.59 \pm 0.15\mu V$)诱发的 N1 成分显著大于中央区($M \pm SE = 0.35 \pm 0.17\mu V$)，$p = 0.021$，95%$CI$ = [0.03, 0.45]，且显著大于右半球($M \pm SE = 0.07 \pm 0.16\mu V$)，$p < 0.001$，95%$CI$ = [0.26, 0.79]；中央区($M \pm SE = 0.35 \pm 0.17\mu V$)诱发的 N1 成分显著大于右半球($M \pm SE = 0.07 \pm 0.16\mu V$)，$p < 0.001$，95%$CI$ = [0.14, 0.43]。组别与脑半球的交互作用显著，$F(2, 23) = 10.90$，$p < 0.001$，$\eta_p^2 = 0.49$。在关注道德标准条件下，左半球($M \pm SE = 0.48 \pm 0.23\mu V$)诱发的 N1 成分显著大于中央区($M \pm SE = 0.13 \pm 0.27\mu V$)，$p = 0.015$，95%$CI$ = [0.06, 0.64]，且显著大于右半球($M \pm SE = -0.08 \pm 0.28\mu V$)，$p = 0.004$，95%$CI$ = [0.16, 0.95]；中央区($M \pm SE = 0.13 \pm 0.27\mu V$)诱发的 N1 成分显著大于右半球($M \pm SE = -0.08 \pm 0.28\mu V$)，$p = 0.038$，95%$CI$ = [0.01, 0.40]。而在控制条件下，左半球($M \pm SE = 0.70 \pm 0.19\mu V$)诱发的 N1 成分显著大于右半球($M \pm SE = 0.21 \pm 0.17\mu V$)，$p = 0.001$，95%$CI$ = [0.18, 0.80]。中央区($M \pm SE = 0.57 \pm 0.18\mu V$)诱发的 N1 成分显著大于右半球($M \pm SE = 0.21 \pm 0.17\mu V$)，$p < 0.001$，95%$CI$ = [0.19, 0.54]。其他的条件下, 均为无显著差异，$p > 0.05$（见图 11）。

（2）P2(100~200ms)

通过对 P2 成分进行 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：欺骗 vs. 诚实）×5（脑区：额区 vs. 额中区 vs. 中央区 vs. 中顶区 vs. 顶区）的重复测量方差分析可知，组别的主效应不显著性，$F(1, 24) = 0.52$，$p = 0.479$。行为决策的主效应不显著，$F(1, 24) = 0.07$，$p = 0.798$。脑区的主效应不显著，$F(4, 96) = 2.39$，$p = 0.056$。组别与行为决策的交互作用显著，$F(1, 24) = 4.93$，$p = 0.036$，$\eta_p^2 = 0.17$。进一步简单效应分析发现，在关注道德标准条件下，欺骗试次诱发的 P2 成分($M \pm SE$ =1.16 ± 0.44μV)显著小于诚实试次($M \pm SE$= 1.86 ± 0.38μV)，$p = 0.045$，95%CI = [0.015,1.409]；在控制条件下，欺骗试次诱发的 P2 成分($M \pm SE$ =2.22 ± 0.40μV)与诚实试次诱发的 P2 成分($M \pm SE$= 1.42 ± 0.27μV)无显著差异，$p = 0.073$，95%CI = [−1.67, 0.08]；脑区与行为决策的交互作用显著，$F(4, 96) = 3.25$，$p = 0.015$，$\eta_p^2 = 0.12$。进一步简单效应分析发现，在顶区，欺骗试次诱发的 P2 成分($M \pm SE$= 1.62 ± 0.38μV)显著大于诚实试次诱发的 P2 成分($M \pm SE$= 1.03 ± 0.32μV)，$p = 0.049$，95%CI = [0.044,0.745]。其他的条件下，均为无显著差异，$p > 0.05$（见图 11）。

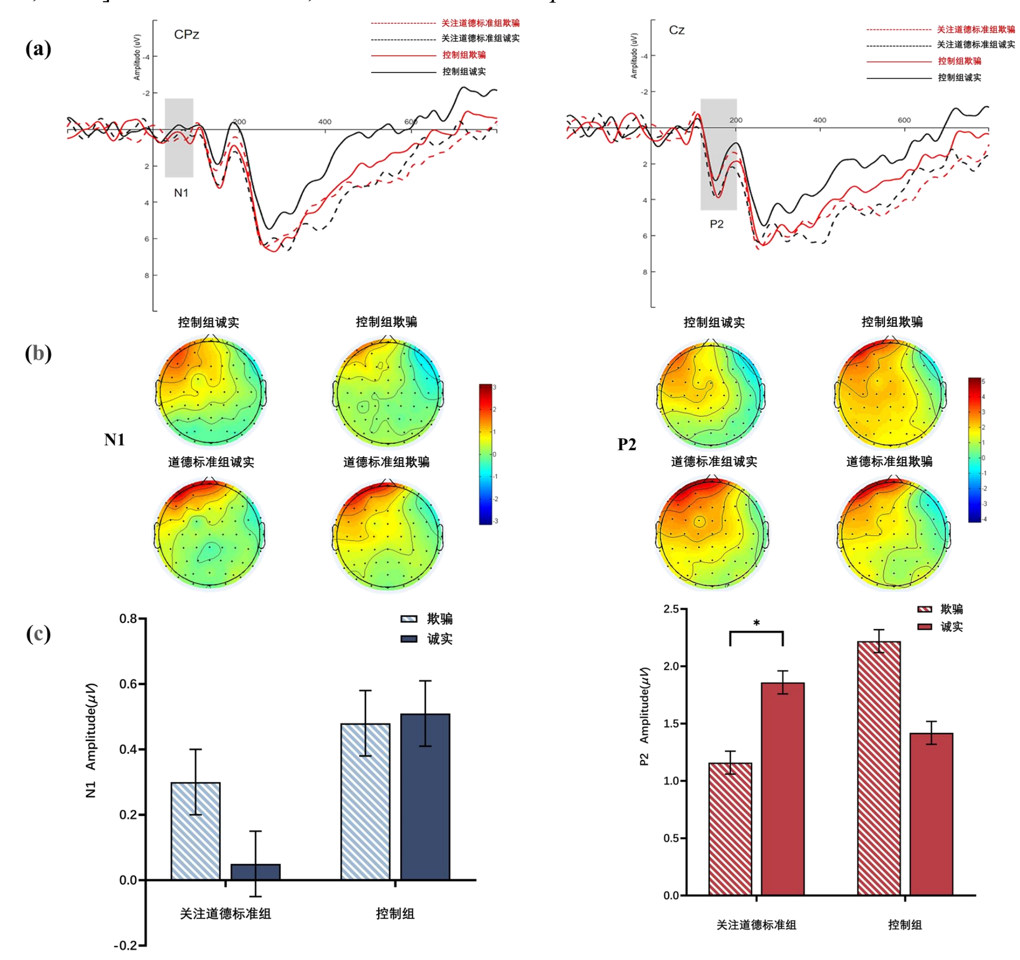


图 11 (a) CPz、Cz 在不同行为决策条件下的平均 ERPs，CPz 图中灰色条表示 N1(20~80ms)的时间窗，Cz 图中灰色条表示 P2 的时间窗(100~200ms)。(b)每种条件下 N1、P2 的脑地形图。(c) 柱状图显示了不同条件下的平均 N1 和 P2 值，误差条表示平均值的标准误差。

通过对 P2 成分进行 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：欺骗 vs. 诚实）×3（脑半球：左半球 vs. 中央区 vs. 右半球）的重复测量方差分析可知，组别的主效应不显著，$F(1, 24) = 0.52$，$p = 0.479$。行为决策的主效应不显著，$F(1, 24) = 0.07$，$p = 0.798$。脑半球的主效应具有显著性的差异，$F(2, 48) = 19.24$，$p < 0.001$，$\eta_p^2 = 0.445$。通过事后比较发现，左半球($M \pm SE = 2.16 \pm 0.28\mu V$)诱发的 P2 成分显著大于右半球($M \pm SE = 0.99 \pm 0.25\mu V$)，$p < 0.001$，95%$CI$ = [0.57, 1.76]；中央区($M \pm SE = 1.84 \pm 0.28\mu V$)诱发的 P2 成分显著大于右半球($M \pm SE = 0.99 \pm 0.25\mu V$)，$p < 0.001$，95%$CI$ = [0.50, 1.20]。组别与行为决策的交互作用显著，$F(1, 24) = 4.93$，$p = 0.036$，$\eta_p^2 = 0.17$。脑半球与组别的交互作用显著，$F(2, 48) = 5.70$，$p = 0.010$，$\eta_p^2 = 0.33$。进一步简单效应分析发现，在控制组中，左半球($M \pm SE = 2.25 \pm 0.29\mu V$)诱发的 P2 成分显著大于右半球($M \pm SE = 1.14 \pm 0.27\mu V$)，$p < 0.001$，95%$CI$ = [0.68, 1.53]；中央区($M \pm SE = 2.07 \pm 0.30\mu V$)诱发的 P2 成分显著大于右半球($M \pm SE = 1.14 \pm 0.27\mu V$)，$p < 0.001$，95%$CI$ = [0.57, 1.28]。在关注道德标准组中，左半球($M \pm SE = 2.07 \pm 0.38\mu V$)诱发的 P2 成分显著大于右半球($M \pm SE = 0.85 \pm 0.40\mu V$)，$p = 0.002$，95%$CI$ = [0.41, 2.05]；中央区($M \pm SE = 1.62 \pm 0.42\mu V$)诱发的 P2 成分显著大于右半球($M \pm SE = 0.85 \pm 0.40\mu V$)，$p < 0.001$，95%$CI$ = [0.38, 1.17]。其他的条件下, 均为无显著差异，$p > 0.05$（见图 11）。

（3）N2(150~250ms)

通过对 N2 成分进行 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：欺骗 vs. 诚实）×5（脑区：额区 vs. 额中区 vs. 中央区 vs. 中顶区 vs. 顶区）的重复测量方差分析可知，组别的主效应不显著，$F(1, 24) = 0.52$，$p = 0.892$。行为决策的主效应不显著，$F(1, 24) = 1.91$，$p = 0.179$。脑区的主效应显著，$F(4, 96) = 13.34$，$p < 0.001$，$\eta_p^2 = 0.36$。通过事后比较发现，额区($M \pm SE = 3.26 \pm 0.50\mu V$)诱发的 N2 成分显著大于顶区($M \pm SE = 1.40 \pm 0.30\mu V$)，$p = 0.023$，95%$CI$ = [0.17, 3.55]；额中区($M \pm SE = 3.32 \pm 0.46\mu V$)诱发的 N2 成分显著大于中顶区($M \pm SE = 2.16 \pm 0.28\mu V$)，$p = 0.004$，95%$CI$ = [0.29, 2.05]，且显著大于顶区($M \pm SE = 1.40 \pm 0.30\mu V$)，$p = 0.004$，95%$CI$ = [0.47, 3.38]。中央区($M \pm SE = 2.94 \pm 0.37\mu V$)诱发的 N2 成分显著大于中顶区($M \pm SE = 2.16 \pm 0.28\mu V$)，$p < 0.001$，95%$CI$ = [0.30, 1.28]，且显著大于顶区($M \pm SE = 1.40 \pm 0.30\mu V$)，$p = 0.002$，95%$CI$ = [0.46, 2.63]。中顶区($M \pm SE = 2.16 \pm 0.28\mu V$)诱发的 P2 成分显著大于顶区($M \pm SE = 1.40 \pm 0.30\mu V$)，$p = 0.012$，95%$CI$ = [−1.40, −0.12]。组别与行为决策的交互作用显著，$F(1, 24) = 6.52$，$p = 0.017$，$\eta_p^2 = 0.21$。进一步简单分析发现，在控制组中，欺骗试次诱发的 N2 成分($M \pm SE = 3.26 \pm 0.43\mu V$)显著大于诚实试次诱发的 P2 成分($M \pm SE = 2.03 \pm 0.40\mu V$)，$p = 0.017$，95%$CI$ = [0.24, 2.21]。在关注道德标准组中，

欺骗试次诱发的 N2 成分($M \pm SE$= 2.31 ± 0.47μV)与诚实试次诱发的 N2 成分($M \pm SE$= 2.87 ± 0.48μV)差异不显著，p = 0.135，95%CI = [−0.19, 1.31]。其他的条件下，均为无显著差异，p > 0.05（见图 12）。

通过对 N2 成分进行 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：欺骗 vs. 诚实）×3（脑半球：左半球 vs. 中央区 vs. 右半球）的重复测量方差分析可知，组别的主效应不显著，F (1, 24) = 0.02，p = 0.892。行为决策的主效应不显著，F (1, 24) = 1.91，p = 0.179。脑半球的主效应具有显著性的差异，F (2, 48) = 9.90，p < 0.001，η_p^2 = 0.29。通过事后比较发现，左半球($M \pm SE$= 2.86 ± 0.36μV)诱发的 N2 成分显著大于右半球($M \pm SE$= 2.01 ± 0.35μV)，p = 0.033，95%CI = [0.06, 1.65]；中央区($M \pm SE$= 2.98 ± 0.38μV)诱发的 N2 成分显著大于右半球($M \pm SE$= 2.01 ± 0.35μV)，p < 0.001，95%CI = [0.51, 1.43]。组别与行为决策的交互作用显著，F (1, 24) = 6.52，p = 0.017，η_p^2 = 0.21，简单分析结果如上（见图 12）。

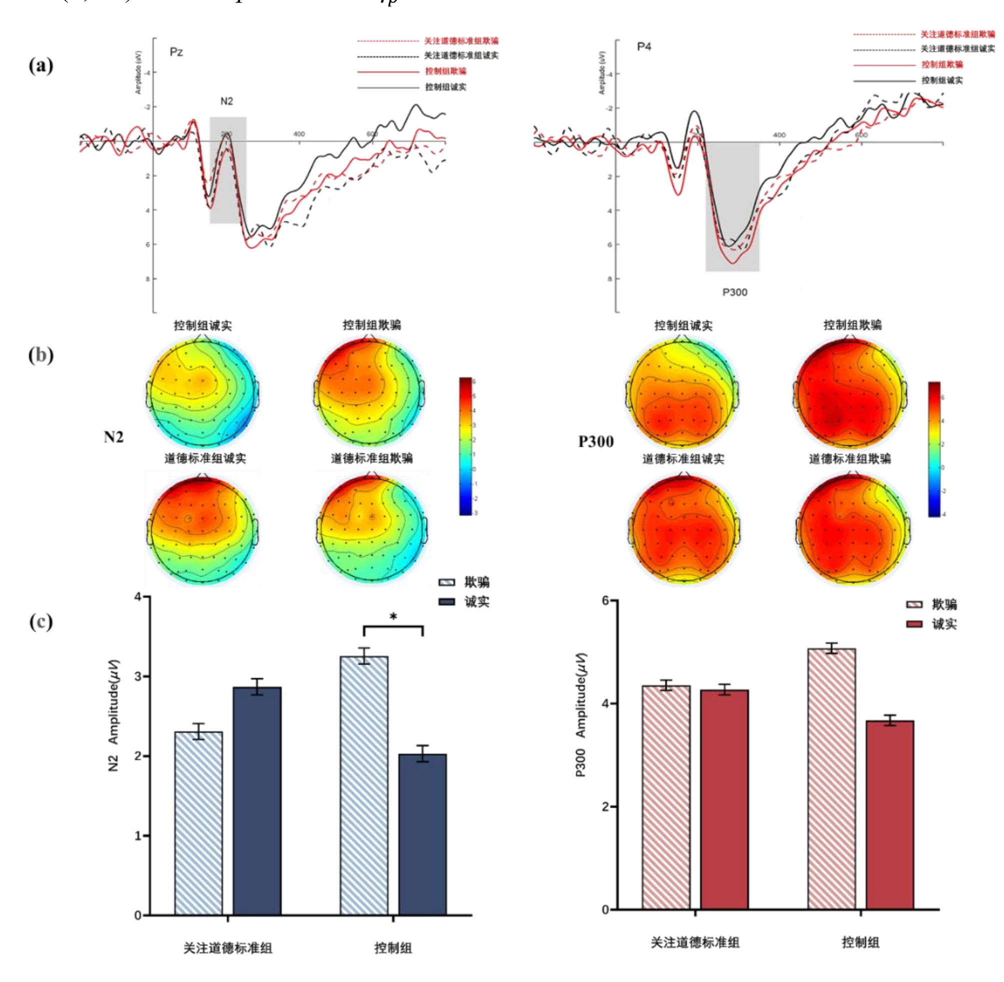


图 12 (a) Pz、P4 在不同行为决策条件下的平均 ERPs，Pz 图中灰色条表示 N2(150~250ms)的时间窗，P4 图中灰色条表示 P300 的时间窗(200~300ms)。(b)每种条件下 N2、P300 的脑地形图。(c)柱状图显示了不同条件下的平均 N2 和 P300 值，误差条表示平均值的标准误差。

（4）P300(200~300ms)

通过对 P300 成分进行 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：欺骗 vs. 诚实）×5（脑区：额区 vs. 额中区 vs. 中央区 vs. 中顶区 vs. 顶区）的重复测量方差分析可知，组别的主效应不显著，$F(1, 24) = 0.02$，$p = 0.879$。行为决策的主效应显著，$F(1, 24) = 5.56$，$p = 0.027$，$\eta_p^2 = 0.19$。通过事后检验发现，相比于诚实试次($M \pm SE$= 3.97 ± 0.50μV)，在欺骗试次($M \pm SE$= 4.72 ± 0.47μV)中诱发了更大的 P300 波幅，$p = 0.027$，95%CI = [0.09,1.39]。其他的条件下, 均为无显著差异，$p > 0.05$（见图 12）。

通过对 P300 成分进行 2（组别：关注道德标准组 vs. 控制组）×2（行为决策：欺骗 vs. 诚实）×3（脑半球：左半球 vs. 中央区 vs. 右半球）的重复测量方差分析可知，组别的主效应不显著，$F(1, 24) = 0.02$，$p = 0.879$。行为决策的主效应显著，$F(1, 24) = 5.56$，$p = 0.027$，$\eta_p^2 = 0.19$，事后检验如上。其他的条件下, 均为无显著差异，$p > 0.05$（见图 12）。

3.4 实验 2 讨论

实验 2 的行为结果发现，在道德标准启动条件下，被试在欺骗试次中选择低于真实信念的比例与诚实试次中相比无显著差异。而在控制条件下，被试在欺骗试次中选择低于真实信念的比例显著大于诚实试次中诚实试次中的比例。这些结果可能表明控制组被试在欺骗行为中有更强的自我欺骗倾向，即通过更悲观的随机概率估计来维护道德自我。而在道德标准启动条件下，自我欺骗倾向不明显，表明提升对道德标准的关注可抑制不道德行为中的自我欺骗(Batson, 1999; Tang et al., 2018) ，验证了假设 2。

脑电结果发现，在控制条件下，欺骗试次诱发自我欺骗的 P2 成分与诚实试次无显著差异。在道德标准启动条件下，欺骗试次诱发自我欺骗的 P2 显著小于诚实试次。这一结果表明，当启动道德标准时，个体在不道德行为中的自我欺骗行为受到了抑制。从电生理机制的角度来看，在道德标准启动条件下，P2 成分的减小可能反映了个体在道德决策中对不道德选项的自动抑制。这种抑制可能与前额叶皮层的激活有关，该区域在道德判断和行为控制中起着关键作用(罗跃嘉 等, 2013)。相比较其他脑区，欺骗试次条件在顶区诱发的 P2 成分显著大于诚实试次。顶区的 P2 成分激活表明了与行为决策过程中的道德冲突处理有关(Abe et al., 2007; Lee et al., 2009)。这表明个体在欺骗试次中的自我欺骗更容易产生认知冲突，可能因为个体在不道德行为中需要处理内在的道德标准与个人利益之间的冲突。左侧大脑的 P2 成分激活更为明显，这与实验 1 的发现相一致，进一步强调了左侧大脑在注意力分配和情绪认知加工中的重要作用(Carretie´ et al., 2001; 王益文, 林崇德, 2005)。在自我欺骗的情况下，个体需要更多的情绪性动机参与。因此，当道德标准被启动时，可能增强了与情绪调节相关的脑区的活动，从而抑制了不道德行为中的自我欺骗(范伟 等, 2022)。脑电结果还发现在道

德标准启动条件下，欺骗试次诱发自我欺骗的 N2 成分与诚实试次无显著差异。在控制条件下，欺骗试次诱发自我欺骗的 N2 显著大于诚实试次，与实验 1 的结果一致。N2 成分是执行控制功能的指标，反映了对冲突的监控与解决(Ofen et al., 2016)。这些研究结果可能表明，关注道德标准可以减弱自我欺骗的心理倾向，因为在关注道德标准组，N2 波幅较小且欺骗条件与诚实条件之间无显著差异。而在控制组中，较大的 N2 波幅则标示了自我欺骗的发生，个体仍利用自我欺骗来合理化行为以缓解内部冲突(Moore, 2016; Tenbrunsel & Messick, 2004)。同时，相比于诚实试次，被试在欺骗试次中的自我欺骗会诱发更大的 P300 波幅，这与实验 1 的结果保持了一致。

4 总讨论

本研究通过发送者-接受者范式诱发被试的不道德行为，并通过个体对未知随机概率值的预测来测量自我欺骗，深入探讨了不道德行为对自我欺骗的影响。随后，研究通过道德标准启动任务，考察道德标准如何调节自我欺骗，并比较了实验组和控制组的行为表现及脑电波幅差异。结果表明，个体在不道德行为中倾向于采取比真实信念更小的虚假信念来维护道德自我，并通过对随机概率的悲观解释合理化其不道德行为。进一步分析发现，提高对道德标准的关注可以抑制不道德行为中的自我欺骗。已有研究表明，自我欺骗是一种用于应对自身利益与道德标准冲突的心理策略。根据道德褪色理论，个体通过自我欺骗淡化其行为的道德内涵，以便合理化不道德行为并维持自我形象。然而，这种心理机制在不道德行为中的普遍性阻碍了个体的道德发展，并可能导致不道德行为的频繁发生(Levy, 2004; Lu & Chang, 2011; Rick et al., 2008; Tenbrunsel et al., 2010; Tenbrunsel & Messick, 2004)。本研究结合了不道德行为范式与自我欺骗范式，深入探讨了自我欺骗在不道德行为中的作用，同时也考察了通过增强道德标准关注度来减少不道德行为中自我欺骗的可能性。

4.1 不道德行为会促进自我欺骗的发生

本研究通过实验 1 和实验 2 的行为结果，验证了不道德行为会促进自我欺骗的发生，验证了假设 1。结果显示，在欺骗试次中，被试选择低于真实信念的比例显著高于诚实试次。这表明，在实施不道德行为时，个体倾向于采取虚假信念以维持其道德自我。具体而言，个体可能通过悲观解读随机概率值，合理化其不道德行为。例如，被试可能通过自我辩解，认为“我欺骗你并不是因为我不道德，而是因为我认为你中奖的概率很低，因此我理应得到更多的奖金”。这一发现扩展了道德褪色理论，进一步揭示了不道德行为中自我欺骗的具体机制(Moore, 2016; Tenbrunsel & Messick, 2004)。自我欺骗的本质在于个体通过利用外部信息，

如未知的随机概率，构建出一种有利于自身的解释框架，从而维持其内在的道德观。这一现象可以通过道德褪色理论解释。根据该理论，个体通过自我欺骗淡化行为的道德含义，使本应具备道德判断的行为重新编码为“非道德性”的选择，最终模糊了行为的伦理维度(Kunda, 1990; Tenbrunsel & Messick, 2004)。本研究中的结果表明，个体会选择性地解读外部线索，并利用这些线索支持自己的道德自我，而非正视其不道德行为的真实性质(Chance & Norton, 2015; Roeser et al., 2016)。已有文献也支持自我欺骗在不道德行为中的重要性。例如，Kirkland(2011)发现，自我欺骗是导致律师职业道德问题的关键因素，个体在面对竞争情境时，往往通过自我欺骗合理化自私行为，导致在道德责任判断和危害估计中产生严重错误(Bok, 1989)。这些结果表明，不道德行为中的自我欺骗不仅会影响个体的道德判断，还可能加剧不道德行为的频发，使得个体更加倾向于通过自我辩解来维护其道德自我 (Moore, 2016; Tenbrunsel & Messick, 2004)。

在实验中，脑电结果表明，与诚实试次相比，在欺骗试次中处于自我欺骗的个体会诱发更大的 P2 成分。这一现象与 P2 成分作为早期判断和视觉刺激加工的指标一致，表明个体在自我欺骗情境中需要更多的认知和情绪性动机参与来进行自我辩解(Carretie´et al., 2001; 范伟 等, 2022)。P2 的增加可能反映了个体在自我欺骗中对不道德行为的合理化需求，通过强化注意力的分配来抵消道德冲突带来的认知不适。此外，相比于诚实试次，在欺骗试次中处于自我欺骗的个体会诱发更大的 N2。这可能反映了自我欺骗的个体在意识层面面临着处理个人利益与道德标准的冲突，需要调动更多的认知资源来进行监控和解决冲突(Ofen et al., 2016)。N2 的增强表明个体在不道德行为中自我欺骗的心理过程，需要通过监控与调节认知冲突来平衡行为与道德之间的不一致性(Moore, 2016; Tenbrunsel & Messick, 2004)。在欺骗过程中，个体可能通过将外部信息解释为支持其行为的证据，以此减少内在的认知冲突，并维持对自我道德形象的正面认知。除了 P2 和 N2，欺骗试次还诱发了更大的 P300 波幅，这不仅反映了执行控制过程的复杂性，还涉及高级认知功能，如决策和记忆 (范伟 等, 2022; Hu et al., 2015)。P300 的增大可能与个体通过自我欺骗减少认知负荷，从而更好地进行欺骗行为有关。这一结果与以往的研究一致，表明 P300 波幅的变化与执行控制功能的需求相关(Yang et al., 2024)。例如，研究发现高认知负荷情境下的个体更倾向于自我欺骗，而自我欺骗可以帮助个体降低在不道德行为中的内在冲突，从而诱发更大的 P300 波幅 (Chen et al., 2014; Jian et al., 2019)。总的来说，本研究的脑电结果表明，自我欺骗在不道德行为中的发生不仅伴随着更大的早期视觉和情绪加工需求（P2 波幅的增加），还需要更多的认知资源来解决道德冲突（N2 波幅的增强），并通过减少认知负荷来支持复杂的决策过程（P300 波

幅的增大），验证了假设 3。这些结果进一步证明了自我欺骗作为应对不道德行为的一种心理策略的重要性。

4.2 提升道德标准的关注可以抑制不道德行为中的自我欺骗

实验 2 的结果表明，在启动道德标准后，个体的自我欺骗倾向显著减弱。行为结果表明，在道德标准启动条件下，被试在欺骗试次中选择低于真实信念的比例与诚实试次相比，两者无显著差异，而在控制条件下，欺骗试次中选择低于真实信念的比例显著大于诚实试次。这一结果表明，控制组的被试在欺骗行为中表现出更强的自我欺骗倾向，即通过更悲观的随机概率估计来维护其道德自我。通过提升道德标准的关注，可以有效减少这种通过自我欺骗来维护道德形象的倾向(Batson, 1999; Tang et al., 2018)，验证了假设 2。

脑电结果进一步支持了这一结论。在控制条件下，欺骗试次诱发自我欺骗的 P2 成分与诚实试次无显著差异。在道德标准启动条件下，欺骗试次诱发自我欺骗的 P2 显著小于诚实试次。P2 与早期注意力加工和情绪认知相关(Carretie´ et al., 2001)，其降低表明在道德标准启动条件下，个体在欺骗行为中不需要额外的情绪动机来维持自我欺骗。另一方面，N2 成分通常与冲突监控和执行控制有关(Ofen et al., 2016)，道德标准启动后，欺骗与诚实试次中的 N2 波幅无显著差异，而在控制条件下，欺骗试次诱发的 N2 显著大于诚实试次。这说明道德标准启动可以减少个体在欺骗行为中面临的内部冲突，削弱自我欺骗所需的认知调节过程(Moore, 2016; Tenbrunsel & Messick, 2004)。这些结果可以通过自我概念维持理论来解释。自我概念维持理论认为，个体在面对不道德行为时会合理化其行为以维持积极的自我形象(Mitchell et al., 1997; Tenbrunsel & Messick, 2004)。然而，当个体被提示关注道德标准时，他们对不道德行为的合理化过程被阻碍，更倾向于严格评估其行为的道德性。这种对道德标准的关注能够促进自我监督，减少自我欺骗倾向，抑制不道德行为的发生(Bandura et al., 1996; Bering et al., 2005)。此外，P300 成分的结果表明，个体在欺骗试次中的自我欺骗诱发的 P300 波幅大于诚实试次，反映了自我欺骗过程中更复杂的认知加工。这与先前研究一致，P300 通常与决策过程中的责任归因和动机性回忆有关(Debey et al., 2012; Hu et al., 2015)。自我欺骗需要个体在认知上更积极地处理不道德行为的后果，以减轻由行为与道德信念不一致带来的情绪冲突(范伟 等, 2022; Farrow, 2015)。相比较其他脑区，欺骗试次诱发的自我欺骗在顶区的 P2 成分显著大于诚实试次，进一步表明了自我欺骗在不道德行为中的作用，该区域负责处理与道德决策相关的冲突评估(Abe et al., 2007; Young et al., 2007)。总之，本研究的结果表明，提升个体对道德标准的关注能够显著抑制不道德行为中的自我欺骗倾向。这为减少自

我欺骗及其负面社会影响提供了新的理论依据和实践路径。

4.3 研究不足与展望

本研究采用 ERPs 技术深入探讨了不道德行为中道德标准对自我欺骗的影响。通过实验研究，我们揭示了道德标准对自我欺骗的抑制作用，尤其是道德标准在不道德行为中通过减少虚假信念的维护和弱化相关认知冲突与情绪动机对自我欺骗的影响。然而，研究仍存在一些局限性和不足之处。首先，实验主要依赖于实验室控制环境中的特定任务，未能完全反映不道德行为在实际生活中的复杂性。未来的研究可以通过更多现实情境中的实验设计，进一步探讨不道德行为对自我欺骗的普遍性及其在不同情境下的表现。其次，本研究的参与者群体仅限于在校大学生，且样本量较为有限。虽然这些参与者在实验中提供了有价值的结果，但不同年龄、社会背景和文化背景的个体可能会在道德判断和自我欺骗的行为上存在差异。因此，未来的研究应当扩大样本群体，以提高研究结论的外部效度。虽然本研究使用了 ERP 技术揭示了相关脑电波幅的变化，但更先进的技术如 fNIRS、fMRI 等可以帮助更精确地揭示自我欺骗的神经机制。因此，未来研究可以采用更高分辨率的神经成像技术，进一步探索自我欺骗在脑内的激活模式和其与道德判断相关的神经基础。

5 结论

本研究通过两个实验探讨了不道德行为中自我欺骗的心理作用及其神经机制，特别是探究道德标准对自我欺骗的抑制作用。实验 1 的结果表明，不道德行为会促进自我欺骗的发生。实验 2 进一步证明，提升对道德标准的关注可以显著减少个体在不道德行为中的自我欺骗倾向，具体表现为个体在欺骗行为中对虚假信念的维护减少，以及相关的认知冲突（如 N2 波幅）和情绪动机（如 P2 波幅）的弱化。这些结果支持自我概念维持理论，说明道德标准的启动通过阻碍合理化过程，有效抑制了自我欺骗的发生。道德标准不仅能促使个体在道德冲突中进行更严格的自我评估，还能减少个体利用自我欺骗维护道德自我的可能性。本研究为理解不道德行为中的自我欺骗机制提供了新的视角，并为未来道德干预措施和教育实践提供了理论依据。

参考文献

Abe, N., Suzuki, M., Mori, E., Itoh, M., & Fujii, T. (2007). Deceiving others: distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions. *Journal of*

Cognitive Neuroscience, *19*(2), 287−295.

Andreoni, J., & Sanchez, A. (2020). Fooling myself or fooling observers? avoiding social pressures by manipulating perceptions of deservingness of others. *Economic Inquiry*, *58*(1), 12−33.

Babino, A., Makse, H. A., DiTella, R., & Sigman, M. (2018). Maintaining trust when agents can engage in self-deception. *Proceedings of the National Academy of Sciences, 115*(35), 872−873.

Bandura, A. (2011). Self-deception: A paradox revisited. *Behavioral and Brain Sciences, 34*(1), 16−17.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, *71*(2), 364−374.

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, *77*(3), 525−537.

Bering, J. M., McLeod, K., & Shackelford, T. K. (2005). Reasoning about dead agents reveals possible adaptive trends. *Human Nature*, *16*, 360−381.

Bok, S. (1985). Secrets and deception: Implications for the military. *Naval War College Review*, *38*(2),73−80.

Boudreau, C., McCubbins, M. D., & Seana, C. (2009). Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. *Social Cognitive and Affective Neuroscience*, *4*(1), 23–34.

Carretie´, L., Mercado, F., Tapia, M., Hinojosa, J. (2001). Emotion, attention, and the "negativity bias", studied through eventrelated potentials. *International Journal of Psychophysiology*, *41*(1), 75–85.

Chance, Z., & Norton, M. I. (2015). The what and why of self-deception. *Current Opinion in Psychology, 6*, 104−107.

Chen, Y., Zhong, Y. P., Zhou, H. B., Zhang, S. M., Tan, Q. B., & Fan, W (2014). Evidence for implicit self-positivity bias: An event–related brain potential study. *Experimental Brain Research*, *232*(3), 985–994.

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2008). Thecontributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex, 19*(7), 1557–1566.

Cui, F., Wu, S., Wu, H., Wang, C., Jiao, C., & Luo, Y. (2017). Altruistic and self-serving goals modulate behavioral and neural responses in deception. *Social Cognitive and Affective Neuroscience*, *13*(1), 63–71.

Cuthbert, B. N., Schupp, H. T., Bradley, M., McManis, M., & Lang, P. J. (1998). Probing affective pictures: Attended startle and tone probes. *Psychophysiology*, *35*(3), 344–347.

Debey, E., Verschuere, B., & Crombez, G. (2012). Lying and executive control: An experimental investigation

using ego depletion and goal neglect. *Acta Psychologica, 140*(2), 133–141.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9−21.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63*(4), 568–584.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology, 79*(6), 861–875.

Epley, N., & Whitchurch, E. (2008). Mirror, mirror on the wall: Enhancement in self-recognition. *Personality and Social Psychology Bulletin, 34*(9), 1159–1170.

Fan, W., Ren, M., Zhang, W., & Zhong, Y. (2022). The impact of feedback on self-deception: Evidence from ERP. *Acta Psychologica Sinica*, *54*(5), 481−496.

[范伟, 任梦梦, 张文洁, 钟毅平. (2022). 反馈对自我欺骗的影响：来自 ERP 的证据. *心理学报*, *54*(5), 481–496.]

Fan, W., Ren, M., Zhang, W., Xiao, P., & Zhong, Y. (2020). Higher self-control, less deception: The effect of self-control on deception behaviors. *Advances in Cognitive Psychology*, *16*(3), 228−241.

Fan, W., Yang, B., Liu, J., & Fu, X. (2017). Self-deception: For adjusting individual psychological states. *Advances in Psychological Science*, *25*(8), 1349–1359.

[范伟, 杨博, 刘娟, 傅小兰. (2017). 自我欺骗：为了调节个体心理状态. *心理科学进展*, *25*(8), 1349–1359.]

Fan, W., Yang, Y., Zhang, W., & Zhong, Y. (2021). Ego Depletion and Time Pressure Promote Spontaneous Deception: An Event-Related Potential Study. *Advances in Cognitive Psychology*, *17*(3), 221−229.

Farrow, Burgess, Wilkinson, & Hunter. (2015). Neural correlates of self-deception and impression-management. *Neuropsychologia, 67*, 159−174.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175−191.

Festinger, & Freedman. (1964). Dissonance reduction and moral values. *Personality Change*, 220−243.

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011−1027.

Foster, G., & Frijters, P. (2014). The formation of expectations: Competing theories and new evidence. *Journal of Behavioral and Experimental Economics*, *53*, 66–81.

Gangl, K., Pfabigan, D. M., Lamm, C., Kirchler, E., & Hofmann, E. (2017). Coercive and legitimate authority

impact tax honesty: evidence from behavioral and ERP experiments. *Social Cognitive and Affective Neuroscience*, *12*(7), 1108–1117.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35*(7), 603−618.

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology, 37*(2), 147−169.

Hirschfeld, R. R., Thomas, C. H., & McNatt, D. B. (2008). Implications of self-deception for self-reported intrinsic and extrinsic motivational dispositions and actual learning performance: A higher order structural model. *Educational and Psychological Measurement*, *68*(1), 154–173.

Hu, X., Pornpattananangkul, N., & Nusslock, R. (2015). Executive control-and reward-related neural processes associated with the opportunity to engage in voluntary dishonest moral decision making. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 475–491.

Jian, Z., Zhang, W., Tian, L., Fan, W., & Zhong, Y. (2019). Self-Deception Reduces Cognitive Load: The Role of Involuntary Conscious Memory Impairment. *Frontiers in Psychology*, *10*, 1718.

Johnson, D. D., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, *477*(7364), 317−320.

Johnson, E. A. (1995). Self-deceptive coping: Adaptive only in ambiguous contexts. *Journal of Personality, 63*(4), 759–791.

Jones. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review, 16*(2), 366–395.

Ju, S., Zhao, Y., & Fu, X.,(2003). A Study of the Mechanism of Self-Deception. *Journal of Sun Yatsen University(Social Science Edition)*, *43*(5), 19–26.

[鞠实儿, 赵艺, & 傅小兰. (2003). 自我欺骗的认知机制. *中山大学学报 (社会科学版)*, *43*(5), 19–26.]

Kirkland, K. (2011). Self-deception and the pursuit of ethical practice: challenges faced by large law firm general counsel. *U. St. Thomas LJ, 9*, 593–618.

Kish-Gephart, J. J., Harrison, D. A., & Treviño, L. K. (2010). Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology, 95*(1), 1–31.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480−498.

Lee, S., & Klein, H. J. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology*, *87*(6), 1175−1182.

Lee, T. M., Au, R. K., Liu, H. L., Ting, K. H., Huang, C. M., & Chan, C. C. (2009). Are errors differentiable from deceptive responses when feigning memory impairment? An fMRI study. *Brain and Cognition*, *69*(2), 406−412.

Levy, N. (2004). Self-deception and moral responsibility. *Ratio, 17*(3), 294–311.

Lu, H. J., & Chang, L. (2014). Deceiving yourself to better deceive high-status compared to equal-status others. *Evolutionary Psychology, 12*(3), 635–654.

Martin, L. E., & Potts, G. F. (2009). Impulsivity in decision-making: An event-related potential investigation. *Personality and Individual Differences*, *46*(3), 303–308.

Martínez-González, J. M., L., R. V., I., & Verdejo-García. (2016). Self-deception as a mechanism for the maintenance of drug addiction. *Psicothema, 28*(1), 13–19.

McGregor, I. (2006). Offensive defensiveness: Toward an integrative neuroscience of compen-satory zeal after mortality salience, personal uncertainty, and other poignant self-threats. *Psychological Inquiry, 17*(4), 299–308.

McGregor, I., Zanna, M. P., Holmes, J. G., & Spencer, S. J. (2001). Compensatory convic-tion in the face of personal uncertainty: Going to extremes and being oneself. *Journal of Personality and Social Psychology, 80*(3), 472–488.

Mischel, W. (1999). Implications of person -situation interaction: Getting over the field′ s borderline personality disorder. *European Journal of Personality, 13*(5), 455–461.

Mitchell, Thompson, Peterson, & Cronc. (1997). Temporal Adjustments in the Evaluation of Events: The "Rosy View". *Journal of Experimental Social Psychology, 33*(4), 421–448.

Moore, C. (2016). Always the hero to ourselves: The role of self-deception in unethical behavior. In J.-W. van Prooijen & P. A. M. van Lange (Eds.), *Cheating, Corruption, and Concealment: the Roots of Dishonesty* (pp. 98–119). Cambridge University Press.

Ofen, N., Whitfield-Gabrieli, S., Chai, X. J., Schwarzlose, R. F., & Gabrieli, J. D. (2016). Neural correlates of deception: lying about past events and personal beliefs. *Social Cognitive and Affective Neuroscience*, *12*(1), 116–127.

Pears, D. F., & Pugmire, D. (1982). Motivated irrationality. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *56*, 157–196.

Pinker, S. (2011). Representations and decision rules in the theory of self-deception. *Behavioral and Brain Sciences*. *34*(1), 35–37.

Pires, L., Leitão, J., Guerrini, C., & Simões, M. R. (2014). Event-related brain potentials in the study of inhibition: cognitive control, source localization and age-related modulations. *Neuropsychology Review*, *24*, 461–490.

Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, *6*, 278.

Politzer, & Carles. (2001). Belief revision and uncertain reasoning. *Thinking & Reasoning, 7*(3), 217–234.

Potts, G.F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition*, *56*(1), 5–13.

Rick, S., Loewenstein, G., Monterosso, J. R., Langleben, D. D., Mazar, N., Amir, O., & Ariely, D. (2008). Commentaries and Rejoinder to" The Dishonesty of Honest People". *Journal of Marketing Research*, *45*(6), 645−653.

Roeser, K., Mc Gregor, V. E., Stegmaier, S., Mathew, J., K, & Meule, A. (2016). The Dark Triad of personality and unethical behavior at different times of day. *Personality and Individual Differences, 88*, 73−77.

Rottenburger, J. R., Carter, C. R., & Kaufmann, L. (2019). It's alright, it's just a bluff: Why do corporate codes reduce lying, but not bluffing?. *Journal of Purchasing and Supply Management*, *25*(1), 30−39.

Samad, Z. (2021). Self-Deception: Adopting False Beliefs for a Favorable Self-View. In *Self-Deception: Adopting False Beliefs for a Favorable Self-View: Samad, Zeeshan*. [Sl]: SSRN.

Schotter, A., & Trevino, I. (2014). Belief elicitation in the laboratory. *Annu. Rev. Econ.*, *6*(1), 103−128.

Shaw, M., Quezada, S. A., & Zárate, M. A. (2011). Violence with a conscience: Religiosity and moral certainty as predictors of support for violent warfare. *Psychology of Violence*, *1*(4), 275.

Sheridan, Z., Boman, P., Mergler, A., & Furlong, M. J. (2015). Examining well-being, anxiety, and self-deception in university students. *Cogent Psychology*, *2*(1), 993850

Shu., & Gino. (2012). Sweeping dishonesty under the rug: how unethical actions lead to forgetting of moral rules. *Journal of Personality and Social Psychology, 102*(6), 1164−1177.

Shuster, & Levy. (2020). Contribution of self-and other-regarding motives to (dis) honesty. *Scientific Reports*, *10*(1), 1−11.

Suchotzki, K., Crombez, G., Smulders, F. T., Meijer, E., & Verschuere, B. (2015). The cognitive mechanisms underlying deception: An event-related potential study. *International Journal of Psychophysiology*, *95*(3), 395−405.

Surbey, M. K. (2011). Adaptive significance of low levels of self-deception and cooperation in depression. *Evolution and Human Behavior, 32*(1), 29−40.

Tang, H., Wang, S., Liang, Z., Sinnott-Armstrong, W., Su, S., & Liu, C. (2018). Are proselfs more deceptive and hypocritical? social image concerns in appearing fair. *Frontiers in Psychology*, *9*, 2268.

Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, *17*(2), 223−236.

Tenbrunsel, Diekmann, Wade-Benzoni, Kimberly, & Bazerman. (2010). The ethical mirage: A temporal explanation as to why we are not as ethical as we think we are. *Research in Organizational Behavior, 30*, 153–173.

Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences, 907*(1), 114−131.

Turk, M. (2012). The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life edited. *Cognitive Neuropsychiatry*, *18*(2),146−151.

Turner. (1975). Moral Weakness, Self-Deception and Self-Knowledge. *New Blackfriars*, *56*(662), 294–305.

Von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, *34*(1), 1−16.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, *20*(1), 28−32.

Wang Y., & Lin, C. (2005). Frontal Involvement to Executive Control: Load Effects Reflected by. *Acta Psychologica Sinica*, *37*(6), 723–728.

[王益文, & 林崇德. (2005). 额叶参与执行控制的 ERP 负荷效应. *心理学报*, *37*(6), 723–728.]

Wu, H., Hu, X., & Fu, G. (2009). Does willingness affect the N2-P3 effect of deceptive and honest responses? *Neuroscience Letters*, *467*(2), 63–66.

Yang, Y., Zhong, B., Zhang, W., & Fan, W. (2024). The impact of social comparison on self-deception: An event-related potentials study. *Cognitive, Affective, & Behavioral Neuroscience*, *24*(5), 931−947.

Young, L., & Koenigs, M. (2007). Investigating emotion in moral cognition: A review of evidence from functional neuroimaging and neuropsychology. *British Medical Bulletin*, *84*(1), 69−79.

Zheltyakova, Kireev, Korotkov, & Medvedev. (2020). Neural mechanisms of deception in a social context: an fMRI replication study. *Scientific Reports*, *10*(1), 1−12

Zhong, L., & M, L. (2019). A commentary on empirical study of self-deception. *Journal of Psychological Science*, *42*(3), 709–714.

[钟罗金, & 莫雷. (2019). 自我欺骗实证研究述评. *心理科学*, *42*(3), 709–714.]

Zhong, L., Ru, T., Fan, M., & Mo, L. (2019). The effect of cognitive vagueness and motivation on conscious and unconscious self-deception. *Acta Psychologica Sinica*, *51*(12), 1330–1340.

[钟罗金, 汝涛涛, 范梦, & 莫雷. (2019). 认知模糊程度和动机强度对有意识和无意识自我欺骗的影响. *心理学报*, *51*(12), 1330–1340.]

The effect of ethical standards on self-deception in unethical behavior: Evidence from ERP

FAN Wei[1,2,3] *YANG Ying*[1,2] *GUO Xiya* [1,2] *LIN Zhuoming* [1,2] *ZHONG Yiping*[1,2,3]

([1] Department of Psychology, Hunan Normal University, Changsha 410081, China)

([2] Cognition and Human Behavior Key Laboratory of Hunan Province, Changsha 410081, China)

([3] Institute of Interdisciplinary Studies, Hunan Normal University, Changsha 410081, China)

Abstract Self-deception refers to an individual's motivated distortion of facts, resulting in false beliefs that contradict true beliefs and deviate from reality. Self-deception is a complex, widespread psychological phenomenon. While research often emphasizes its positive effects, its negative impacts on mental health, behavior, and society—particularly within the moral domain—should not be overlooked. As self-deception is pervasive in immoral behavior, it exacerbates immoral conduct and leads to serious consequences. Therefore, studying the inhibitory effect of moral standards on self-deception is crucial for understanding its broader implications.

This study aims to explore the psychological role and neural mechanisms of self-deception in immoral behavior using event-related potential (ERP) technology, focusing on how moral standards inhibit self-deception. Experiment 1 investigates the neural basis of self-deception in immoral behavior. In this experiment, immoral behavior was induced in participants using the sender-receiver paradigm, and self-deception was measured through participants' predictions of random probability values. Behavioral results revealed that, in deception trials, participants were significantly more likely to make predictions that underestimated their true beliefs compared to honest trials. EEG results showed that, compared to honest trials, deception trials evoked larger N2 and P300 components. Further analysis found that in the centroparietal and parietal regions, deception trials elicited larger P2 components compared to honest trials. Experiment 2 employed a moral standards priming task to investigate how attention to moral standards influences self-deception, aiming to compare behavioral responses and EEG amplitude differences between experimental and control groups. Under control conditions, behavioral results indicated that participants in deception trials were significantly more likely to make predictions that underestimated their true beliefs compared to honest trials. EEG results showed that, under the moral standards priming condition, the P2 and N2 components elicited during deception trials

were significantly lower than those in honest trials. These findings suggest that in immoral behavior, participants are more prone to forming false beliefs, leading to self-deception. Enhanced attention to moral standards can effectively reduce self-deception.

This study explored the psychological role and neural mechanisms of self-deception in immoral behavior through two experiments, focusing on how moral standards inhibit it. Experiment 1 revealed that immoral behavior facilitates self-deception, while Experiment 2 confirmed that increasing attention to moral standards significantly reduces the tendency for self-deception, as shown by reduced false beliefs, cognitive conflict, and emotional motivation. The results support the self-concept maintenance theory, indicating that moral standards effectively inhibit self-deception by interfering with the rationalization process. This study provides valuable insights into the mechanisms of self-deception and suggests novel approaches for moral interventions.

Key words self-deception, immoral behavior, moral standards, beliefs, event-related potentials